

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw.model

Run on: May 29, 2003, 15:14:20 ; Search time 3987 Seconds
(without alignments)
11387.102 Million cell updates/sec

Title: US-09-934-066-1

Perfect score: 1560

Sequence: i ctcaagaatcagattcaa.....gaaaaaaaaaaaaaaaaaaaa 1560

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1540	98.7	1563	8	AV120765	AV120765 Arabidops
2	1518.8	97.4	1541	8	AV084227	AV084227 Arabidops
3	832	53.3	83253	8	AF000383	AF000383 Arabidops
4	627.2	40.2	656	8	AF070375	AF070375 Arabidops
5	467.6	30.0	1576	8	AF238384	AF238384 Vigna rad
6	467.4	30.0	1980	8	BVU309173	BVU309173 Beta vulg
7	466	29.9	1659	8	D89972	D89972 Vigna mungo
8	463	29.7	1839	8	CSCYSPRNS	CSCYSPRNS
9	454.8	29.2	1485	8	AV133531	AV133531 Arabidops
10	454.8	29.2	1516	8	AV059104	AV059104 Arabidops
11	454.8	29.2	1760	8	AF424619	AF424619 Arabidops
12	454.8	29.2	1779	8	AF370160	AF370160 Arabidops
13	451.4	28.9	1850	8	PVZ99956	PVZ99956 Phaseolus v
14	444	28.5	1323	6	E05717	E05717 Asparaginil
15	442.2	28.3	1437	8	AV090296	AV090296 Arabidops
16	442.2	28.3	1616	8	AV058055	AV058055 Arabidops
17	441.6	28.3	1323	6	E05718	E05718 Asparaginil
18	435.6	27.9	1834	8	VSCYSPROB	VSCYSPROB
19	424.2	27.2	1724	8	D89971	D89971 Vigna mungo
20	421.8	27.0	1572	8	AF260827	AF260827 Ipomoea b
21	393.2	25.2	1140	8	VNA23882	VNA23882 Vicia nar
22	386	24.7	1866	8	AB025310	AB025310 Oryza sat
23	385.8	24.7	1919	8	AF082347	AF082347 Zea mays
24	385.2	24.7	1864	8	ZMAL131719	ZMAL131719 Zea mays
25	383	24.6	1529	8	LES243876	LES243876 Lycopersi
26	377.6	24.2	1762	8	RCCVPE	RCCVPE
27	376.2	24.1	1569	8	ZMAL131718	ZMAL131718 Zea mays
28	373	23.9	1840	8	VNCYSTPRO	VNCYSTPRO
29	362.6	23.2	1461	8	AV059156	AV059156 Arabidops
30	362.6	23.2	1643	8	AF367254	AF367254 Arabidops
31	361.6	23.2	1652	8	PVZ99957	PVZ99957 Phaseolus v
32	351	22.5	1736	8	SOYCPA	SOYCPA
33	347.2	22.3	1437	8	NTA238880	NTA238880 Nicotiana
34	345.6	22.2	1836	8	CENAEI	CENAEI
35	345	22.1	1323	6	E05716	E05716 Asparaginil
36	343.4	22.0	1323	6	E05720	E05720 Asparaginil
37	342	21.9	1851	8	VSA007743	VSA007743 Vicia sat
38	338.8	21.7	1715	8	AF169019	AF169019 Glycine m
39	336.4	21.6	1649	8	AF169973	AF169973 Sesamum i
40	314.4	20.2	1152	6	E05721	E05721 Asparaginil
41	306.8	19.7	1360	8	NTA238881	NTA238881 Nicotiana
42	305.4	19.6	894	6	E05722	E05722 Asparaginil
43	304	19.5	1485	8	AF082346	AF082346 Hordeum v
44	302.4	19.4	813	6	E05723	E05723 Asparaginil
45	294	18.8	1642	8	AV062178	AV062178 Oryza sat

ALIGNMENTS

RESULT 1	AV120765	1563 bp	mrna	linear	PLN 22-JUN-2002
LOCUS	AV120765	Arabidopsis thaliana vacuolar processing enzyme/asparaginyl			
DEFINITION	Arabidopsis thaliana vacuolar processing enzyme/asparaginyl	endopeptidase, putative (At3g20210) mRNA, complete cds.			
ACCESSION	AV120765				
VERSION	AV120765.1	GI:21539540			
KEYWORDS	FLI_CDNA.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 1563)				
AUTHORS	Southwick,A., Karlin-Neumann,G., Nguyen,M., Tripp,M., Miranda,M.,				

Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H., Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A., Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K., Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K., Hayashizaki, K., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (11-JUN-2002) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 e-mail for correspondence: arab@sequence.stanford.edu

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: "RIKEN Arabidopsis Full-Length cDNA"): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PDEC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Southwick, A., Nguyen, M., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K., Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Southwick, A. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES
source

Location/Qualifiers
 1. .1563
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="RAFL09-06-A19 (R15622)"
 /note="This clone is in pBluescript
 ecotype: Columbia"

gene

1. .1563
 /gene="At3g20210"

CDS

63. .1463
 /note="MAL21.23"
 /gene="At3g20210"
 /codon_start=1
 /product="vacuolar processing enzyme/asparaginyl
 endopeptidase, putative"
 /protein_id="AA53323.1"
 /db_xref="GI:21539541"
 /translation="MSPGCHQOILVFLHALLIFSAESRKTOLLNDNVESDSKAG
 TRAVLVASNEYNYNRHQNDCHAYQLKRGKLDENITVFNHDDIAFSSNPRPGV
 IIRPGDEYIKGPDYIKREAVNQFNVLGNESGVYGGNKKVYKSGPNDNFIIY
 YADGAPLLIAMPTGDEWAKDFNEVLEHRKRYNKKVYIYVEACSGMPEGLIKK
 NLIYVTAANSKSSWGYCPESYPPSEIGTCLGDTFSISWLESDLDHDMSEKTL
 EQYHYVVRGSDVPETSRVCFGRFEMKDYLLSYIGRNPENDNFYTFESFSPIS
 NSGLVNPRIPLLILQRIKAPMGSLSEKAKKLLDEKHKRKOIDSITDIURLSV
 KQTNVLNLLSTFTTGOPLVDWDCEKTLVNSFKNHCGATVHYLUKYGALANICNMG
 VDVKQTVSAFEQACSN"

BASE COUNT
ORIGIN

508 a 290 c 322 g 443 t

Query Match 98.7%; Score 1540; DB 8; Length 1563;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1540; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCACAGAATCAGATCAAGATAGAAGTTTTCACAAACAATGCTAGTCTCTTGGTCAC 60
 DB 24 CTCACAGAATCAGATCAAGATAGAAGTTTTCACAAACAATGCTAGTCTCTTGGTCAC 83
 QY 61 TTTCAGATCTCTGTTTCTTCATGCTTTCATCTCTCAGCTCAGTCCCGCAAAACC 120
 DB 84 TTTCAGATCTCTGTTTCTTCATGCTTTCATCTCTCAGCTCAGTCCCGCAAAACC 143
 QY 121 CAATTGCTGACCATATGATGTGATCTAGCGACAGAGTGCAGAGGCAAGGACGATGG 180
 DB 144 CAATTGCTGACCATATGATGTGATCTAGCGACAGAGTGCAGAGGCAAGGACGATGG 203

181 GCTGTTTGTAGTTGCTGGATCAAAATGAATATTATACTACAGGCATCAGGCTGACATATGC 240
 DB 204 GCTGTTTGTAGTTGCTGGATCAAAATGAATATTATACTACAGGCATCAGGCTGACATATGC 263
 QY 241 CACGCGTATCAGATCTCTCCGAAAAGCGGTTTAAAAGATGAAAACATCATCTGTTTATG 300
 DB 264 CACGCGTATCAGATCTCTCCGAAAAGCGGTTTAAAAGATGAAAACATCATCTGTTTATG 323
 QY 301 TATGATGATATCGGTTTCTCGGAGAAATCCTTAGGCCCTGGAGTTATCATTAATAACCA 360
 DB 324 TATGATGATATCGGTTTCTCGGAGAAATCCTTAGGCCCTGGAGTTATCATTAATAACCA 383
 QY 361 GATGGAGAAAGATGTTTATAAGGAGTTCCTTAGGACTACACTAAAGAGCTCTTAATGTT 420
 DB 384 GATGGAGAAAGTGTATTATAAGGAGTTCCTTAGGACTACACTAAAGAGCTCTTAATGTT 443
 QY 421 CAAACTTCTACAAATGTGTTTACTTGGAAATGAAGTGGGCTCACAGGAGGAATGGCAAA 480
 DB 444 CAAACTTCTACAAATGTGTTTACTTGGAAATGAAGTGGGCTCACAGGAGGAATGGCAAA 503
 QY 481 GTTGTGAAAAGTGGTCTTAATGATAATATCTTTCATCTATTATGCTGACCATGGAGTCT 540
 DB 504 GTTGTGAAAAGTGGTCTTAATGATAATATCTTTCATCTATTATGCTGACCATGGAGTCT 563
 QY 541 GGCTTAATACGATGCCACTGCTGATGAAGTTATGGCAAAAGATTTCAATGAAGTCTTGT 600
 DB 564 GGCTTAATACGATGCCACTGCTGATGAAGTTATGGCAAAAGATTTCAATGAAGTCTTGT 623
 QY 601 GAGAAGATGCATAAGAGAGAAAATAACAAAGATGCTGATCTATGTTGAAGCATGTGAA 660
 DB 624 GAGAAGATGCATAAGAGAGAAAATAACAAAGATGCTGATCTATGTTGAAGCATGTGAA 683
 QY 661 TCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAATCTCAACATATACGACATGACTGCT 720
 DB 684 TCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAATCTCAACATATACGACATGACTGCT 743
 QY 721 GCTAAATCTAAAGAGAGCAGCTGGGAGTTTACTGCTCCTGAGTACATATCTCCTCCTCT 780
 DB 744 GCTAAATCTAAAGAGAGCAGCTGGGAGTTTACTGCTCCTGAGTACATATCTCCTCCTCT 803
 QY 781 TCTCAGATTCGAACTTCTCTCGCGGATACATTTAGCATCTCTTGGCTTCAGGACAGTAC 840
 DB 804 TCTCAGATTCGAACTTCTCTCGCGGATACATTTAGCATCTCTTGGCTTCAGGACAGTAC 863
 QY 841 CTTTCATGACATGAGCAAAAGAGACTTTTGGAGCAACAATACCATGTTGTTAAAGAGAGTA 900
 DB 864 CTTTCATGACATGAGCAAAAGAGACTTTTGGAGCAACAATACCATGTTGTTAAAGAGAGTA 923
 QY 901 GGATCTGATGTACAGAGACTTCTCATGTATGCGGTTTCGGACACAGAGAGATGCTTAA 960
 DB 924 GGATCTGATGTACAGAGACTTCTCATGTATGCGGTTTCGGACACAGAGAGATGCTTAA 983
 QY 961 GATTATCTTTCTCTTTACATTTGGAAGAAATTCCTGAAAACGATAACTTCACTTTTCACG 1020
 DB 984 GATTATCTTTCTCTTTACATTTGGAAGAAATTCCTGAAAACGATAACTTCACTTTTCACG 1043
 QY 1021 TCCCTTTTCTCTCAGCAATCTCTAAATCTTGGCTTGGTCAATCCGCGGATATTCCTCTGCTA 1080
 DB 1044 TCCCTTTTCTCTCAGCAATCTCTAAATCTTGGCTTGGTCAATCCGCGGATATTCCTCTGCTA 1103
 QY 1081 TACTCTCAGAGAAAGATTTCAAAAAGCTCCCAATGGGATCACTTTGAAAAGCAAAAGCTCAG 1140
 DB 1104 TACTCTCAGAGAAAGATTTCAAAAAGCTCCCAATGGGATCACTTTGAAAAGCAAAAGCTCAG 1163
 QY 1141 AAGAAATGCTTCCACGAAAGAGATCATAGGAACAATCGATCAGAGCAATTAACAGACATT 1200
 DB 1164 AAGAAATGCTTCCACGAAAGAGATCATAGGAACAATCGATCAGAGCAATTAACAGACATT 1223
 QY 1201 CTGCGGCTTACAGTTTAAACCAACCAATGCTTAAATCTCTTAACATCCAGACACAACA 1260
 DB 1224 CTGCGGCTTACAGTTTAAACCAACCAATGCTTAAATCTCTTAACATCCAGACACAACA 1283
 QY 1261 GGACAGCCTCTTGTAGACGATTCGGGATTCGTTCAAGACTCTAGTTAATAGCTTCAAGANT 1320

Db 1284 GGACAGCCTCTGTAGACGATGGGATGCTTCAAGACTCTAGTAAATAGCTTCAAGAAAT 1343
Qy 1321 CACTGCGGTGCAACGGTGCATTACGGATTGAAGTATACAGGAGCGCTTGCCAAATATCTGC 1380
Db 1344 CACTGCGGTGCAACGGTGCATTACGGATTGAAGTATACAGGAGCGCTTGCCAAATATCTGC 1403
Qy 1381 AATATGGAGTGGATGGAAGCAAACTGTGTCAGCCATTGAACAGCTGTGTCAGATGAA 1440
Db 1404 AATATGGAGTGGATGGAAGCAAACTGTGTCAGCCATTGAACAGCTGTGTCAGATGAA 1463
Qy 1441 TGATTTGCAAAACAAATGTGATATTCGACTTTTAAATAATCAAACTTAAATTCATATAAAC 1500
Db 1464 TGATTTGCAAAACAAATGTGATATTCGACTTTTAAATAATCAAACTTAAATTCATATAAAC 1523
Qy 1501 TCGATGTAGAGATGGTGGTTCATGATGACTACTTTTACAT 1540
Db 1524 TCGATGTAGAGATGGTGGTTCATGATGACTACTTTTACAT 1563

RESULT 2
LOCUS AY084227 1541 bp mRNA linear PLN 21-JUN-2002
DEFINITION Arabidopsis thaliana clone 100372 mRNA, complete sequence.
ACCESSION AY084227
VERSION AY084227.1 GI:21402937
KEYWORDS FLI_CDNA.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1541)
Haas,B.J., Volfovsky,N., Town,C.D., Troupkhan,M., Alexandrov,N.,
Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
Genome Biol. (2002) In press
2 (bases 1 to 1541)
Brover,V., Troupkhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 1541)
Brover,V., Troupkhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the Ws or Laer ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Genet carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.
FEATURES
Location/Qualifiers
1..1541
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="100372"
59..1459
/codon_start=1
/product="vacuolar processing enzyme/asparaginyl

endopeptidase, putative"
/protein_id="AAMG0827.1"
/db_xref="GI:21536495"
/translation="MSSPLGHFOILVFLHLLIFSABSRKTLTOLLNDVSSDSKSAKG
TRAVLVAGSNENYNYRHQADICHAYOILRGGGLKDELILVFMVDDIAFVSENPBGV
IINKPDGEDYKGVPKDYTKAVNVQNFYNVLLNGESVGTGNGKVKVSGPNDNFYI
YADGAPGLIAMPTGDEVMARDFNEVLEIKHRRKIYNKMYIYVACSGSGMFGILKK
NLNIAYVATANSKSSWGVCPEYPPPEISIGTCLGDTFSISLESDULHMSKETL
EQOYHVYKRVGSDVPETSHVCKMLKDYLSYIGRNPNDNFTFESFSPIS
NSGLVNPDRDPLLYLORKIOKAPMGSLSEKAOKKLLDEKNHRKOIQDSITDIIRLSV
KOTVNLNLLSTRITGLOPLVDWDFCTVLNVSFNHCGATVHYGLKYTGALANICMG
VDVKQTSATEOAGSM"
BASE COUNT 499 a 287 c 321 g 434 t
ORIGIN
Query Match 97.4%; Score 1518.8; DB 8; Length 1541;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1520; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CTCACAAGATCAGATCAAGATGAAGATTTTCAACAATGTCTAGTCCTCTTGGTCAC 60
Db 20 CTCACAAGATCAGATCAAGATGAAGATTTTCAACAATGTCTAGTCCTCTTGGTCAC 79
Qy 61 TTTTCAGATTTCTTTTCTTTCATGCTTTCCTTCTCAGCTGAGTCCCGCAAAACC 120
Db 80 TTTTCAGATTTCTTTTCTTTCATGCTTTCCTTCTCAGCTGAGTCCCGCAAAACC 139
Qy 121 CAATTGCTGAACGATATGATGTTGAATCTAGCGACAAAGAGTGCAGGAGGACACGATGG 180
Db 140 CAATTGCTGAACGATATGATGTTGAATCTAGCGACAAAGAGTGCAGGAGGACACGATGG 199
Qy 181 GCTGTTTGTAGTTCGTCGATCAAAATGAATATATTAACACTACAGGATCAGGCTGACATATGC 240
Db 200 GCTGTTTGTAGTTCGTCGATCAAAATGAATATATTAACACTACAGGATCAGGCTGACATATGC 259
Qy 241 CACGCGTATCAGATACCTCCGAAAGGCGGTTTAAAGATGAAGACATCATTTGTGTTATG 300
Db 260 CACGCGTATCAGATACCTCCGAAAGGCGGTTTAAAGATGAAGACATCATTTGTGTTATG 319
Qy 301 TATGATGATATCGCGTTCCTTCGAGAAATCCTAGGCGCTGGAGTTATCATTTAAATAACCA 360
Db 320 TATGATGATATCGCGTTCCTTCGAGAAATCCTAGGCGCTGGAGTTATCATTTAAATAACCA 379
Qy 361 GATGAGAGATGTTTATAAAGAGATTCCTTAAGGACTACACTAAAGAGCTGTTAATGTT 420
Db 380 GATGAGAGATGTTTATAAAGAGATTCCTTAAGGACTACACTAAAGAGCTGTTAATGTT 439
Qy 421 CAAAACITCTACATGTTTACTTGGAAATGAAAGTGGCGTCCACAGGAGGAAATGGCAAA 480
Db 440 CAAAACITCTACATGTTTACTTGGAAATGAAAGTGGCGTCCACAGGAGGAAATGGCAAA 499
Qy 481 GTTGTGAAAAGTGGTTCCTTAATGATAATATCTTCATCTATTATGCTGACCATGGAGCTCT 540
Db 500 GTTGTGAAAAGTGGTTCCTTAATGATAATATCTTCATCTATTATGCTGACCATGGAGCTCT 559
Qy 541 GGCTTAATAGGATGCCCGACCTGGTGAAGATTTATGGCAAAAGATTTCAATGAAGCTTTG 600
Db 560 GGCTTAATAGGATGCCCGACCTGGTGAAGATTTATGGCAAAAGATTTCAATGAAGCTTTG 619
Qy 601 GAGAGATGCATAGAGAAAAAATACAAAGATGGTGATCTATGTTGAAGCATGTGAA 660
Db 620 GAGAGATGCATAGAGAAAAAATATACAAAGATGGTGATCTATGTTGAAGCATGTGAA 679
Qy 661 TCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTCAACATATACCGAGTGAAGTCT 720
Db 680 TCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTCAACATATACCGAGTGAAGTCT 739
Qy 721 GCTAATTTCTAAAGAGAGCAGCTGGGAGATTTTACTGTCCTGAGTATATCTCTCTCTCT 780
Db 740 GCTAATTTCTAAAGAGAGCAGCTGGGAGATTTTACTGTCCTGAGTATATCTCTCTCTCT 799
Qy 781 TCTGAGATTTGGAAGTCTGTCGCGGATACATTTAGCATCTCTTGGCTTTGAGGAGTAC 840

```

Db      800  TCTGAGATGGAACTTCTCTCGCGGATACATATTAGCATCTCTTGGCTTGAGCAGAGTGAC 859
QY      841  CTTTCATGACATGACAAAGAGACTTTGGAGCAACAATACACAGCTTCTTAAGAGAGAGAGTA 900
DB      860  CTTTCATGACATGACAAAGAGACTTTGGAGCAACAATACACAGCTTCTTAAGAGAGAGAGTA 919
QY      901  GGATCTGATGATGACAGAGACTTCTCATGTATGCGGTTTCGGACAGAGAGATGCTTAA 960
DB      920  GGATCTGATGATGACAGAGACTTCTCATGTATGCGGTTTCGGACAGAGAGATGCTTAA 979
QY      961  GATTATCTTCTCTTACATTTGAAGAAATCCTGAAACGATACATTTCACTTTTCAGGAA 1020
DB      980  GATTATCTTCTCTTACATTTGAAGAAATCCTGAAACGATACATTTCACTTTTCAGGAA 1039
QY      1021  TCCCTTTTCCACCAATCTCTAATCTCGGCTTCTGATCGGCGGATATCTCTGCTA 1080
DB      1040  TCCCTTTTCCACCAATCTCTAATCTCGGCTTCTGATCGGCGGATATCTCTGCTA 1099
QY      1081  TACCTCCAGAGAAAGATTCAAAAGCTCCCAATGGATCACTTGAAGCAAGAAAGCTCAG 1140
DB      1100  TACCTCCAGAGAAAGATTCAAAAGCTCCCAATGGATCACTTGAAGCAAGAAAGCTCAG 1159
QY      1141  AAGAAATGCTTGACGAAAGATCATAGAAACAATCGATCAGAGCAATTTACAGACAT 1200
DB      1160  AAGAAATGCTTGACGAAAGATCATAGAAACAATCGATCAGAGCAATTTACAGACAT 1219
QY      1201  CTGCGGCTTTCAGTTAAACAAACAATGCTTAAATCTCTTAACTTCCACAAAGAAACA 1260
DB      1220  CTGCGGCTTTCAGTTAAACAAACAATGCTTAAATCTCTTAACTTCCACAAAGAAACA 1279
QY      1261  GGACAGCCTCTTGTAGACGATTGGGATGCTTCAAGACTCTAGTTAATAGCTTCAAGAT 1320
DB      1280  GGACAGCCTCTTGTAGACGATTGGGATGCTTCAAGACTCTAGTTAATAGCTTCAAGAT 1339
QY      1321  CACTCGGCTCAACGGTGCATTACGGATTGAAGTATACAGAGCGCTTCCCAATATCTGC 1380
DB      1340  CACTCGGCTCAACGGTGCATTACGGATTGAAGTATACAGAGCGCTTCCCAATATCTGC 1399
QY      1381  AATATGGAGTGGATGTGAAGCAAACTGTTTCAGCACTTGAACAAGCTTGTTCGATGTA 1440
DB      1400  AATATGGAGTGGATGTGAAGCAAACTGTTTCAGCACTTGAACAAGCTTGTTCGATGTA 1459
QY      1441  TGATTTCCAAACAATCTGATATTCACACTTTAAATAATCAAGTTAATTTCAATAAACA 1500
DB      1460  TGATTTCCAAACAATCTGATATTCACACTTTAAATAATCAAGTTAATTTCAATAAACA 1519
QY      1501  TCGATGTAGAGATGGTGGTTC 1522
DB      1520  TCGATGTAGAGATGGTGGTTC 1541

RESULT 3
AP000383
LOCUS      8253 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone:MAL21.
ACCESSION AP000383 BA000014
VERSION    AP000383.1 GI:5672520
KEYWORDS
SOURCE
  Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
  clone:MAL21.
ORGANISM
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
  1 (sites)
REFERENCE
  Kaneko,T., Kato,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.
  Structural analysis of Arabidopsis thaliana chromosome 3 II.
  Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
  and BAC clones
  DNA Res. 7 (3), 217-221 (2000)
  20363099
REFERENCE
  Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  (bases 1 to 83253)
  2
  AUTHORS
  TITLE
  JOURNAL
  MEDLINE
  REFERENCE
  AUTHORS

```

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yane,
Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
Tel: 81-438-52-3935, Fax: 81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd/graph.cgi?c=MAL21
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
product similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremli.ni.zooi.fasstate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MZ19 and the 3' clone is MQ12.
Location/Qualifiers
1..83253
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MAL21"
/clone_lib="Mitsui P1"
complement(1..376)
/product="nucleoid chloroplast DNA-binding protein-like"
/note="CDS is reported in Acc# AP002050
gene_id:MZ19.7"
/number=1
/evidence=not_experimental
complement(join(4273..4315,4392..4510,4727..4902,
5273..5277,5836..5914,6019..6096,6222..6360,6555..6620,
7231..7266,7418..7486,7659..7724,7870..8163))
/note="gene_id:MAL21.2"
/codon_start=1
/evidence=not_experimental
/product="protein arginine N-methyltransferase-like
protein"
/protein_id="BAB01859.1"
/db_xref="GI:9293936"
/translation="MQSGDFNSGFGHRELELEDKOGPSLSFGRKRSKSHAGAR
DPRGLANLVRSQDLGEHKSLETSPPCTDFDVAFHSAHVGIHEEMIKDRAR
TETRYEAIHQHSLIEGKVVVDVCGTGILSIFCAQAGAKRVYAVDASJAVQAKREV
KANGLSKDVTLHGRVEDVEIDEDVDIISNMGYMLLYSEMLSVITARDRLKPGV
LILPESHATLYMAPISHPDYSHSIDFWRNVYIDMSAMQAKQAFEPSPESISGE
NVLTWPEVFPLEFGSPAPAKTSETSIASGSSISIPSGEVNKKRTNPDSALVLS
TSPSPPTTHQQTIVFYDPIDVEDQDIEGVTLSQSKENKRFNHNHLEYSACKRSF
VKESVMR"
10472..11680
/note="emb|CA878723.1
gene_id:MAL21.3
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01860.1"
/db_xref="GI:9293937"
/translation="MTMMSDLSQDLLEETLSBVPRTSLCAVRSCKRWNTLFKDRILC
KAEETDQFQFIMKKYKLSMSRFDLNGTLENGGTEFVDPFSKELCHFFNQVKISKVF
QCGLLLCVTKEDINIRLVNPNYGLQIRWIESGNTNIRLFDYALGIDNNRKHILRF
LEYINFMKHFLEIETDFSSNRVLVDIAPRWETESYQFGASLKGNFTYFAKRIE
YEEDGFEPPADNLLCTEYDFSGQLPLFPQHYLDVGLUSSLDKLAALFOCGD
TDUSEVEIVTTLTETNTVSNMPLFKVDMEPHYGRSFMFDYGGCFSSVYVPSLVQINOI
OFDESMTRYVEDATYIIGENGYYKVRLEGEAFANOGGCGYFPCVCFSSVYVPSLVQINOI
VGPKKEREE"

FEATURES
source

exon

CDS

CDS

CDS	join(15535..15806,15005..16155,16253..16435,16674..16748,16945..17100,17229..17327,17425..17552,17741..17819,17917..18282) /note="gene_id: MAL21.4" /codon_start=1 /evidence=not_experimental /product="chloroplast outer envelope hexokinase" /protein_id="BAB01861.1" /db_xref="GI:9293958" /translation="MGVLYMLTAAAVVACSVATVVRMRGRKRWVRVGLLKDLEACETPLGRLOWDAIVEMOAGLVSEGGSKLMLLTFVDDLPNGSETTYVALHLGSSYFRIKVLHGORSLEQVDVERHSIPTSLMNSTSEVLFDFLASSLOFTIEKEGNDFSLSQPLKRELATTFPPYKQTSISSCYLLKWTGFAISEMAGEDIAECLOQALNKRGLDRLVNDTVALSGHFDPTTAAVFGTSGNACYLERTDALIKQNPRTTSGVMVNMWGNFSRPLRTSYDLEDAESNMDSMGFEKMGIMGDIIVRVILRMSEQSDIFGPTSSILSTFVLRTNSVAMHEDDTSELOEARILKDLGVSEYPMKVRKLVKICDVVTRRAALAAAGIAGILKVGCRDGGGRRSDKOIMRRTVVAVEGGLYLNRMFREYDEALRDILGEDVAOVVVVKAMEDGSSIGSALLASSOSVOTIPSV" complement(join(18762..18845,18933..19038,19118..19221,19306..19434,19520..19624,19780..19914,19994..20080,20208..20304,20350..20767,20863..20928,21116..21211,21364..21414,21501..21560,21669..21755,21868..21957,22170..22234,22412..22484)) /note="gene_id: MAL21.5" /codon_start=1 /evidence=not_experimental /product="chaperonin, t-complex protein alpha subunit" /protein_id="BAB01862.1" /db_xref="GI:9293959" /translation="MSTSAONPDISGDRSGODVTONVMCAQAVSNIVKTSLSGPVGLDKMLVDIGDVTINDGATILRMLEVEHPARAKVIELAQDREVGGDTTSVIVAAELKLRANDLVNRNKIITSLISGRVLIAMRESCKYIEELKVTVEKILGVPGLINCAKTSMSKLTSGDSDFANLVAVKAVSNKQNGEIKYPIKGINILKAHQSGARDSYLLNGYALNTRAAQMPFLVSPAKIACLDENLQKTMQLGVQVVVNDPRLKIRQREADMTKE RIEKLKAGANVITTKIGIDMALYFEAGAIARVRKEDMRHVAKATCATLVTTFADMEGETOPAHILSGADEVVEERADDDVILIKGKTSSAVSLLRGANDYMIDEME RALHDALCIKVTLESNTVAGGAVSALSYLEHLATILGSRQELAIAGAFDALLI IPKVLAANAKADTELVAKLRAHYHTAQTKADKKHYSRKKRCSMGLDVLNGTIRNNLEAGVIEPAMSKVITIQFATEAAITILRIDDMIKLVKDESGEE" complement(join(23145..23191,23289..23348,23451..23514,23623..23850,23934..23961,24383..24501)) /note="gene_id: MAL21.6" /codon_start=1 /evidence=not_experimental /product="ubiquitin conjugating protein-like" /protein_id="BAB01863.1" /db_xref="GI:9293960" /translation="MATVNGYTGNTPAATTPAATGSKOSAPPTKTVDSSHVLKPLQSELMGLMGADPGISAFPEEDNIFCWKGTITGSKDVPFGTEYRLSLTFSNDYPPFKSPRVKFETCCHPNVDLIGNICLDILQDKWSAYDVRTILLISLLEGPNISPLNQAAQLWSNQEEYRKMVEKLYPLNA" complement(join(25168..25358,25476..25644,25764..25981,26173..26234,26335..26425,26620..26689,26796..26843)) /note="gene_id: MAL21.7" unknown protein /codon_start=1 /evidence=not_experimental /protein_id="BAB01864.1" /db_xref="GI:9293961" /translation="MEALYAKLYDKYTKLOKKKYSEYDEINKEOEFKPLTVSASEELMEHLRGENOSLEWVKLRNELIISGRDDKFLKCOLKLMEEELKNKSISEEVVCLK ELVGEHPRNVEDSGKKRKRKPESARVTRSMKRSRUSEDLNTVDMPDVSPIKSHH KAKREPLVSPQCCRTTYDGGSSASCTFQALGKHLGMLKSLDNNKMRACIVASHPT TGLSFSFTFINNNGESELLYKPAISLGTQFVAPENWREVIKFTSMCPIFFERVSRTIKLNC" 27734..27973 /note="contains similarity to cytochrome P450 gene_id: MAL21.8" /codon_start=1 /evidence=not_experimental /protein_id="BAB01865.1" /db_xref="GI:9293962" /translation="MMLQCFDWKTKGDKVDNEEAIRGLNLTMAHPLKCIPIVARNLNL	CDS	ACSNLKPKNFEFAKKLLVRGSSAPLMPFPCPSLPV" join(29031..29951,30031..30681) /note="gene_id: MAL21.9" /codon_start=1 /evidence=not_experimental /product="cytochrome P450-like protein" /protein_id="BAB01866.1" /db_xref="GI:9293963" /translation="MAAMNIVDFQNCFFVLLSLFSLCYSLFFRKPKPELRHYDLPPSPPTPIIHLHLLSVLVRSLOTSTKYGSLVLRVSPFVALVSSASTAYEIFREHDVNSISRGPPPTDLSLFAFSFSTAPYGYDMFKMLLVNLLGPOLARSFGFRADELFLVENDLAKAMKESVDICVEALKSNNSICKMIGMSCSEENGEAEVRALATQDGLTKRILLUANRAGFKKLVSLFRKEMDSRDELLERLTVHEDKLDHMHQGTDLALLAACCRDNKAEYKTSRNHIKFFADLLFASDTFTVQTTWTVARIINNPNYLERLGEDSVVSGKARLTQETDLNPLYLQAVVKEGLRHPGLFARFSQEGRRIGG FVRSCKTTLINAVMRSDSWEDPDEPKPERFLASSRSEOKERRRQAIKYATFGS GYRSCPENLAYIFLGAIGVMVQGFWEKRIEKEKVNMEANVGLTWAYLKVTPVPRTLVLPTQNLNRSS" join(32158..32475,32844..33086,33190..33813) /note="contains similarity to cytochrome P450 gene_id: MAL21.10" /codon_start=1 /evidence=not_experimental /protein_id="BAB01867.1" /db_xref="GI:9293964" /translation="MAAMINVEFQNCIFIMLLCLFSLCYSLFFRKPSRQNSQDCDLPPSPSPMPIIGHLHLISLVLHKSQKISSKYGKPLLLHRIENVPILVSSASVAYDIF RVHDKFFLAATLRLLEKIGLSLFOKEIMGVSRGFDELLERILREHEKPDHEHDTDM DALLAAYNDEKAEYKTRNQIKAFIVDMFIAGTDISALTTOGTMAEILNPNVLYRI REKIDSVVGKSRLLQETDLPLKPLYSQAVVKEGLRHPPTLPMVREFQGGCKVGYIIP ASSTLVNNGYAVMRDPNWEPEEFKPERFLASSRSEERIEQALKYIAFGSGRRG CPGTNVYIFVGTAIMGMVQCFDKIRGDKVDMKEAIGGLNLTLAHPLKCTPVARSPP"	CDS	join(34864..34912,34957..35241,35310..35443,35870..36133) Query Match 53.3%; Score 832; DB 8; Length 83253; Best Local Similarity 70.6%; Pred. No. 1.4e-178; Matches 1545; Conservative 0; Mismatches 0; Indels 643; Gaps 7; Qy 1 CTCACAAATCAGATTCACAGATGAGAGTCTTCAACAATGTCTAGTCCCTCTTGGTCCAC 60 Db 72661 CTCACAAATCAGATTCACAGATGAGAGTCTTCAACAATGTCTAGTCCCTCTTGGTCCAC 72720 Qy 61 TTTCCAGATTCCTGTTTCTTCTATGCTTGTGCTTCTTCTCAGCTGAGTCCCGCAAAACC 120 Db 72721 TTTCCAGATTCCTGTTTCTTCTATGCTTGTGCTTCTTCTCAGCTGAGTCCCGCAAAACC 72780 Qy 121 CAATTCGTGACGATAATGATGTTGAATCTAGCCACACAGAGTGCNAAAGGCACACGATGG 180 Db 72781 CAATTCGTGACGATAATGATGTTGAATCTAGCCACACAGAGTGCNAAAGGCACACGATGG 72840 Qy 181 GCTGCTTTAGTTGCTGGATCAAAATGAATATTTAATTAACACAGCAT-----225 Db 72841 GCTGCTTTAGTTGCTGGATCAAAATGAATATTTAATTAACACAGCATCGGTTGTTAAATTA 72900 Qy 226 -----225 Db 72901 TGTTTGAACGTTTAAACATAACAAAAAAGGTCCCAACGAGATTTGTATGAACATA 72960 Qy 226 -----CAGCTGACATATGCCACGCGTATCAGATACTCGGAAAA 264 Db 72961 ATCGACCGACGTTTATTTTCACAGCGTGACATATGCCACGCGTATCAGATACTCGGAAAA 73020 Qy 265 GGCGGTTTAAAAAGATGAAACATCATCTGTTTATGATGATATATCGCGTTTCTCTCG 324 Db 73021 GGCGGTTTAAAAAGATGAAACATCATCTGTTTATGATGATATATCGCGTTTCTCTCG 73080 Qy 325 GAGAACTCCTAGGCGCTGAGTTTATCATTAAATAAACACAGATGAGAGAGTGTATATAAGGA 384 Db 73081 GAGAACTCCTAGGCGCTGAGTTTATCATTAAATAAACACAGATGAGAGAGTGTATATAAGGA 73140 Qy 385 GTTCCTA-----391
-----	--	-----	---	-----	---

Db 73141 GTTCCTAAGGTTCTTATTTCTACTCTTTTGTGGGTTATTTCTACGTTGAATTCATATAC 73200
QY 392 -----AGGACTACACTAAGAAGCTGTGA 415
Db 73201 ATATATATATCAAGTTTGTGTATTTGTTGGGTAGGACTACACTAAGAAGCTGTGA 73260
QY 416 ATGTTCAAAAGCTTCTACAATGTGTACTTGGAAATGAAGTGGCGTCACAGGAGGAAATG 475
Db 73261 ATGTTCAAAAGCTTCTACAATGTGTACTTGGAAATGAAGTGGCGTCACAGGAGGAAATG 73320
QY 476 GCAAAGTTGTGAAAGTGGCTTAAATGATAATATCTTCAATCTATATATGCTGACCATGGAG 535
Db 73321 GCAAAGTTGTGAAAGTGGCTTAAATGATAATATCTTCAATCTATATATGCTGACCATGGAG 73380
QY 536 CTCCTGGCTTAAT ----- 548
Db 73381 CTCCTGGCTTAATAGGTTTCTTAAATTTATGAAATTTATACGTACCATCAATCCATATC 73440
QY 549 -----AGCGATGCCACATG 562
Db 73441 TATAATAAAGATTTTCTCTTGATACACGAACCGGATTTTCTCAGCGATGCCACATG 73500
QY 563 GTGATGAAGTTATGCCAAAAGATTTCAATCAAGTCTTGGAGAAGATGCATAGAGAAAA 622
Db 73501 GTGATGAAGTTATGCCAAAAGATTTCAATCAAGTCTTGGAGAAGATGCATAGAGAAAA 73560
QY 623 AATCACACAGAT ----- 635
Db 73561 AATCACACAGATGGTATATACTCAACCATTCGTTACCCTAGCTTTATACATATGTGTTTC 73620
QY 636 -----GGTGATCTATGTTGAAGCATG 656
Db 73621 TGTTTTTGAATCTATGGTGTGTTTTTTGGATGTTTGGTGATCTATGTTGAAGCATG 73680
QY 657 TGAATCAGGAGATGTTTGAAGGGATTTTAAAGAAAAATCTCAACATATATCGCAGTGAC 716
Db 73681 TGAATCAGGAGATGTTTGAAGGGATTTTAAAGAAAAATCTCAACATATATCGCAGTGAC 73740
QY 717 TGCTGCTAAATCTTAAAGAGAGCAGCTGGGAGTTTACTGTCCTGAGTCATATCTCCTCC 776
Db 73741 TGCTGCTAAATCTTAAAGAGAGCAGCTGGGAGTTTACTGTCCTGAGTCATATCTCCTCC 73800
QY 777 TCCTTCTGAGATTTGGAATCTGCTCGCGGATACATTTAGCATCTCTTGCTGTA ----- 830
Db 73801 TCCTTCTGAGATTTGGAATCTGCTCGCGGATACATTTAGCATCTCTTGCTGTAAGACAG 73860
QY 831 ----- 830
Db 73861 GTACTGCAACAAAAAGATTCAATCCTTATGACTATTGCAATGATTTGATTTGTTCTT 73920
QY 831 -----GGACAGTGAACCTTCATG 847
Db 73921 GAAGAATATTTGTTCAATTTCTTATGTTTTGTGTGTTTGGGACAGTGAACCTTCATG 73980
QY 848 ACATGAGCAAGAGACTTTTGAGCAACAATACACCGTTGTAAGAGAGAGTAGGATCTG 907
Db 73981 ACATGAGCAAGAGACTTTTGAGCAACAATACACCGTTGTAAGAGAGAGTAGGATCTG 74040
QY 908 ATGTACCAGAGACTTCTCATGTATGCCGTTTCCGGAACAGAGAGATGCTTAAAGATTATC 967
Db 74041 ATGTACCAGAGACTTCTCATGTATGCCGTTTCCGGAACAGAGAGATGCTTAAAGATTATC 74100
QY 968 TTTCTCTTACATTTGGAAGAAATCCTGAAACGATAACTTCACTTTACCGGAATCCTTTT 1027
Db 74101 TTTCTCTTACATTTGGAAGAAATCCTGAAACGATAACTTCACTTTACCGGAATCCTTTT 74160
QY 1028 CCTCACCATCTCTTAATCTGCTGCTGCTCAATCCGCGGATATTCCTCTGCTATACCTCC 1087
Db 74161 CCTCACCATCTCTTAATCTGCTGCTGCTCAATCCGCGGATATTCCTCTGCTATACCTCC 74220
QY 1088 AGAGAA ----- 1093
Db 74221 AGAGAAAGGTGAGCTTTTTCGGGTTTTTGTATCATTTTAAACGAAGAGTTTTTCAGCAT 74280

QY 1094 -----AGATTTCAAAAAGCTCCAAATGGATCACTTTGAAAAGCAA 1130
Db 74281 GTTTTAAATGTTTATTCATCTCTTTAGATTTCAAAAAGCTCCAAATGGATCACTTTGAAAAGCAA 74340
QY 1131 AGAAGCTCAGAAGAAATTTGTTGACGAAAAGATCATAGGAAAACAAATCGATCAGAGCAT 1190
Db 74341 AGAAGCTCAGAAGAAATTTGTTGACGAAAAGATCATAGGAAAACAAATCGATCAGAGCAT 74400
QY 1191 TACAGACATTTGCGGGCTTTTCAGTTTAAACAAACCAATGCTTAAATCTCTTTAACTTTCCAC 1250
Db 74401 TACAGACATTTGCGGGCTTTTCAGTTTAAACAAACCAATGCTTAAATCTCTTTAACTTTCCAC 74460
QY 1251 AAGAACACAGCAGACGCTCTTGTAGACGATTTGGGATTTGCTTCAAGACTCTA ----- 1302
Db 74461 AAGAACACAGCAGACGCTCTTGTAGACGATTTGGGATTTGCTTCAAGACTCTAGTAAACAA 74520
QY 1303 ----- 1302
Db 74521 CCACATCTCAACCTTTGTTACTTGTCTTACGCAACCAACCATTTGCAATTTACTATNACC 74580
QY 1303 -----GTTAATAGCTTCAAGAATCACTGGGGTGCACACGGT 1337
Db 74581 AGTGTATATCGAATGAAANTCGCAGGTTAATAGCTTCAAGAAATCACTGGGGTGCACACGGT 74640
QY 1338 GCATTACGGATTTGAAGTATACAGAGCGCTTGGCAATATCTGCAATATGAGGAGTGGATGT 1397
Db 74641 GCATTACGGATTTGAAGTATACAGAGCGCTTGGCAATATCTGCAATATGAGGAGTGGATGT 74700
QY 1398 GAAGCAAACTGTTTTCAGCCATTTGAACAAAGCTTGTTCGATTAATGATTTGCAAAAACATG 1457
Db 74701 GAAGCAAACTGTTTTCAGCCATTTGAACAAAGCTTGTTCGATTAATGATTTGCAAAAACATG 74760
QY 1458 TGATATTCGACTTTTAAATATCAAACTTAAATTTCAATAAACTCGATGTAGAGATGGTT 1517
Db 74761 TGATATTCGACTTTTAAATATCAAACTTAAATTTCAATAAACTCGATGTAGAGATGGTT 74820
QY 1518 GTTTCATGATACTACTTTTACATGAAA 1545
Db 74821 GGTTCATGATACTACTTTTACATGAAA 74848

RESULT 4
AY070375 556 bp mRNA linear PLN 22-APR-2002
LOCUS Arabidopsis thaliana At3g20210 mRNA sequence.
DEFINITION Arabidopsis thaliana
ACCESSION AY070375
VERSION AY070375.1 GI:17979293
KEYWORDS FLI_CDNA.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 656)
Yamada,K., Banh,J., Chang,C.H., Chang,E., Dale,J.M.,
Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.,
Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Yu,S., Bowser,L.,
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
Sakurai,T., Satou,M., Seki,M., Shino,P., Southwick,A.,
Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Arabidopsis Full Length cDNA Clones
Unpublished
2 (bases 1 to 656)
Yamada,K., Banh,J., Chang,C.H., Chang,E., Dale,J.M.,
Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.,
Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Yu,S., Bowser,L.,
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
Sakurai,T., Satou,M., Seki,M., Shino,P., Southwick,A.,

Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission
Submitted (18-DEC-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada, K., Banth, J.,
Chang, C.H., Chang, E., Dale, J.M., Goldsmith, A.D., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C.,
Yamanura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T.,
Karlin-Neumann, G., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M.,
Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Davis, R.W.,
Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGECC)
contributed equally to this work as PIs.

FEATURES
Source
1..656
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RAFL09-59-E07 (R19019)"
/note="This clone is in a modified pBluescript vector
(FLC-1) as a BamHI/XhoI insert.
ecotype: Columbia"
gene
1..656
/gene="At3g20210"
misc_feature
63..647
/gene="At3g20210"
/note="This is a partial cDNA containing no stop codon.
However, it does have a polyA tail which is not present in
the genomic sequence. This cDNA is a partial
correspondence to gene At3g20210."
/evidence="experimental"
BASE COUNT 221 a 113 c 135 g 187 t
ORIGIN
Query Match 40.2%; Score 627.2; DB 8; Length 656;
Best Local Similarity 99.5%; Pred. No. 6.e-132;
Matches 629; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 CTCACAGAATCAGATTCAGATAGAGTGTTCACAAATGCTAGTCTCTTGGTTCAC 60
Db 24 CTCACAGAATCAGATTCAGATAGAGTGTTCACAAATGCTAGTCTCTTGGTTCAC 83
Qy 61 TTTCAGATCTTGTGTTTCTTCATGCTTGTGCTTATCTTCAGTCTGAGTCCCGCAAAACC 120
Db 84 TTTCAGATCTTGTGTTTCTTCATGCTTGTGCTTATCTTCAGTCTGAGTCCCGCAAAACC 143
Qy 121 CAATGCTGCAAGATATGATGTTGAATCTAGGACAAAGTGCACAAAGGCACACGATGG 180
Db 144 CAATGCTGCAAGATATGATGTTGAATCTAGGACAAAGTGCACAAAGGCACACGATGG 203
Qy 181 GCTGTTTTAGTCTGATCAATGAATATATTAACCTACAGGATCAGGCTGACATATGC 240
Db 204 GCTGTTTTAGTCTGATCAATGAATATATTAACCTACAGGATCAGGCTGACATATGC 263
Qy 241 CAGCGGTATCAGATCTCCGAAAGGCGGTTTAAAGATGAAACATCATTTGTTTATG 300
Db 264 CAGCGGTATCAGATCTCCGAAAGGCGGTTTAAAGATGAAACATCATTTGTTTATG 323
Qy 301 TATGATGATATCGGTTTCTCGGAGAAATCTTAGGCTCGAGTTATCATTAATAACCA 360
Db 324 TATGATGATATCGGTTTCTCGGAGAAATCTTAGGCTCGAGTTATCATTAATAACCA 383
Qy 361 GATGGAGAAGATGTTTATAAGAGTTCCTTAAGGACTACACTAAAGAGCTGTATGTT 420

384 GATGGAGAAGATGTTTATAAAGAGGATTCCTTAAGGACTACTAAAGAGCTGTTAATGTT 443
Qy 421 CAAAACCTTCTACAAATGTTTACTTGGAAATGAAAGTGGCTCACAGGAGGAAATGGCAA 480
Db 444 CAAAACCTTCTACAAATGTTTACTTGGAAATGAAAGTGGCTCACAGGAGGAAATGGCAA 503
Qy 481 GTTGTGAAAAGTGGTCTCTAATGATATATCTTCAATGATGATGATGATGATGATGATG 540
Db 504 GTTGTGAAAAGTGGTCTCTAATGATATATCTTCAATGATGATGATGATGATGATGATG 563
Qy 541 GGCTTAAATAGCGATGCCACTGCTGATGAAGTATGCAAAAGATTTTCAATGAAGTCTTG 600
Db 564 GGCTTAAATAGCGATGCCACTGCTGATGAAGTATGCAAAAGATTTTCAATGAAGTCTTG 623
Qy 601 GAGAAGATGATGAAG 632
Db 624 GAGAAGATGATGAAG 655

RESULT 5
AF238384
LOCUS
Vigna radiata asparaginyl endopeptidase (PEL) mRNA, complete cds.
DEFINITION
AF238384
ACCESSION
AF238384.1 GI:13183094
VERSION
KEYWORDS
SOURCE
ORGANISM
Vigna radiata.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Vigna.
REFERENCE
1. (bases 1 to 1576)
AUTHORS
Yang, W.H. and Chen, C.S.
TITLE
Vigna radiata mRNA for asparaginyl endopeptidase, complete cds
JOURNAL
Unpublished
REFERENCE
2. (bases 1 to 1576)
AUTHORS
Yang, W.H. and Chen, C.S.
TITLE
Direct Submission
JOURNAL
Submitted (24-FEB-2000) Institute of Botany, Academia, Sinica,
Nankang, Taipei, Taiwan 11529, Republic of China
FEATURES
source
1..1576
/organism="Vigna radiata"
/cultivar="Tainan 5"
/db_xref="taxon:157791"
/tissue_type="cotyledon"
/dev_stage="germinating seeds"
gene
1..1576
/gene="PEL"
CDS
19..1470
/gene="PEL"
/codon_start=1
/product="asparaginyl endopeptidase"
/protein_id="AAK15049.1"
/db_xref="GI:13183095"
/translation="MATTTTSLTFLFLATVALVAAGRDLYGDFLRLPFSDGNDN
VKGTWAILFAGSNQWYVYRHQADICHAIVLIRKGLKENIIVFVYDDIAFNMDNPR
PGVIINKPDGDVYEGVPKDYTGEDATAHNFYSALLGDKSALTGGSGKVVNSGPDRI
FTYSDHGGPGVLGTPAGPYIYADLVLEVLKKHASKTYKNLVYLEACEAGSIFEL
LPEDINIYATASAEESWGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
ESLHQQYKVKVDRFTLSGGVYGVSHVQYGVDFSKDLFLYLGTDLPANDNLTFVDENSL
WSSSTAVNRDADLVHFWHKFRKAPGSPKKNKARQVLEVMHSHRHIDSVKLVLK
LFGFEKAPVLNAPRAGSALVDDMACLKTMTVTFETHCGSLSOYGMKHMRSFANICN
VGKKQMAEASQAQCVTPASSWSLSLQFSA"

BASE COUNT 419 a 327 c 377 g 453 t
ORIGIN
Query Match 30.0%; Score 467.6; DB 8; Length 1576;
Best Local Similarity 60.8%; Pred. No. 1.e-95;
Matches 780; Conservative 0; Mismatches 499; Indels 3; Gaps 1;
Qy 150 TAGCGACAAGAGTGCAAAAGGCACACGATGGGCTGTTTGTAGTGTGCGATCAATGAATA 209

[illegible]

Db	1215	TCCTGAAGTACTGANTGCTCTTTAGACCGGCTGGATCGGCACCTTGTGTGATGACACTGGCGCTG	1274
Qy	1290	CTTCAAGACTCTAGTTAATAGCTTCAAGAACTACTCGGGTGCACCGTGCATTACGGATT	1349
Db	1275	CTGTAAACCATGGTAGGACTTTTGGAGACAAATTCGGGATCTCTTGTCTCAGTATGGGAT	1334
Qy	1350	GAAGTATACAGGAGCGCTTCCCAATATCTCCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGT	1409
Db	1335	GAACACATAGGCTCTTGTGCAACATCTGCAATCTAGGATAGGAAATAAGAAAGAGCAAAATGGC	1394
Qy	1410	TTACGCCATTGAACAAGCTTGT	1431
Db	1395	TGAAGCCTCAGCACAGCTTGT	1416
RESULT 6			
BVU309173			
LOCUS			
Beta vulgaris mRNA 1980 bp mRNA linear PLN 23-MAY-2002			
DEFINITION			
Accession			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Beta vulgaris			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
Caryophyllales; Caryophyllaceae; Chenopodiaceae; Beta.			
REFERENCE			
AUTHORS			
TITLE			
Kloos, D.U., Oltmanns, H., Dock, C., Stahl, D. and Hehl, R.			
Isolation and molecular analysis of six taproot expressed genes			
J. Exp. Bot. 53 (373), 1533-1534 (2002)			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Submitted (05-APR-2001) Hehl R., Institute of genetics, Technical			
University Braunschweig, Spielmannstr. 7, Braunschweig, D-38106,			
GERMANY			
FEATURES			
source			
CDS			
Location/Qualifiers			
1..1980			
/organism="Beta vulgaris"			
/db_xref="taxon:161934"			
/country="Germany"			
44..1504			
/codon_start=1			
/product="putative vacuolar processing enzyme"			
/protein_id="CAC43295.1"			
/db_xref="GI:14594819"			
/db_xref="SPTREMBL:Q949L7"			
/translation="MMIRVSTGVILVLCVLMSSVDPRLVDNLIIRWSPDHPHSIFPDS			
DVSQVPRWAVLIAGSSGYNRYHQADVCAYOVLLKGLKNDNIIVFYDDDIAYDRDN			
PRGVLTINSPPGHVDYAGVPKDYTGEDVTNNFFPAALIGNKDAITCGSGKTVNVSQ			
HIEFVLDIGGAGLCMPYTPYLADLETLIEKHSQTYKSLVYVTEACSGSLIFE			
GILPEGLNLYATFASNAVSSMTQYCPGODPNVPPEYDTGLDLYSVSLTEQSERLNL			
HTSELQOQVYVETAKTAPRYGVSHVQYQDKELTQDMLYLYGNTNPNNVTVYVDN			
SLHPTSSNAVORDAQLIEFWKFKASQCSQRIINAQKFMEVMSRHWLDDSLIKLI			
SKLLFTEKGLGVLYVRYTCQDLYVDNMLKTLVYTFEKHCSLSQYCKHNRHSIRLI			
ICNAGTITNMAEASACAFSPFSVSLHRRGSA"			
BASE COUNT			
ORIGIN			
558 a 369 c 455 g 598 t			
Query Match			
Best Local Similarity 30.08; Score 467.4; DB 8; Length 1980;			
Matches 774; Conservative 0; Mismatches 511; Indels 0; Gaps 0;			
Qy	151	AGCGACAAGAGTGC AAAAGACACAGATGGCGTGTGTTTATGTTCTGGATCAAAATGATAT	210
Db	170	AGCGATGATGATTTCACTAGGAGCATAGATGGCGGTGTTCTGATTTCCGGATCTAGTGCTTAC	229
Qy	211	TATAACTACAGGCATCAGGTGACATATGCCACCGCTATCATGATCTCCGAAAAGCGGT	270

.Db	230	TGGAAATTACAGGCATCAGGCAGATGTTTGTCTGCTTACCAAGTGTCTAAAAAGGAGGT	289
Qy	271	TTAAAGATGAAACATCATTTGTCTTTATGTATGATGATATCGCGTTTTCCTCGGAGAAT	330
Db	290	CTAAAGGATGAAACATCATTTGTTCATCTGATGATGACATTCGTTATGACGAAGAGAAT	349
Qy	331	CTTAGGCTCGAGTTATCATTAATAAACAGATGGAGAAGATGTTTATAAAGGAGTTCCCT	390
Db	350	CCAAGACCTCGAGTTCTCATTAATAGTCCCTATGGACATGATGTTATGCGAGGATACCA	409
Qy	391	AAGGATACACTAAAGAAGCTGTTAATGTTTCAAAAGTCTTACAATGTGTTTACTTTGGAAT	450
Db	410	AAGGATTAACAGGGGAAGATGTGACTGTCAATAACTTCTTCGCAGCTATACTTTGGAAC	469
Qy	451	GAAGTGGCCTCACAGGAGAAATGGCAAAAGTTCTGAAAAAGTGGTCTTAATGATAATATC	510
Db	470	AAAGATGCCATTACTGTTGCGACGGAAAGTGGTGAATAGCGGACCTAATGATCACATC	529
Qy	511	TTCACTAATTATGCTTGACCATGGAGCTCCCTGGGCTTTAATAGCGATGCCCATCGTGATGA	570
Db	530	TTCACTCTTTTACTCAGACCACTGGTGGTGGTCTGGTGTGCTCGGTATGCTACTATCCATC	589
Qy	571	GTTATGCGAAAGATTTCAATGAAGTCTTCGAGAAGATGCATAGAGAAAAAATACAAC	630
Db	590	CTTTACGCTGATGAGCTCATTTGAACCTTTGAAGGAAAAACATCGCTGCGAACCTATAAA	649
Qy	631	AAGATGGTGTATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAG	690
Db	650	AGCCTGGTAGTCTATATTGAAGCTGTGATCTGGGAGCATCTTTGAGGATATCTTCCT	709
Qy	691	AAAAATCTCAACATATACGCAGTGCATGCTGCTTAATTTCTAAAGAGCAGCTGGGAGTT	750
Db	710	GAAGGACTTAAATATATATGCTACCACTGCTTCAATGCTCTAGAGAGTAGTTGGGAACC	769
Qy	751	TACTGTCTGAGTCATATCTCTCTCTCTCTCTGAGATTGGAACCTGTCTCGCGGATACA	810
Db	770	TATTTGCTCGACAGGATCTTAATGTTTCTCCAGAGTATGATACCTGTTTGGGTGATTG	829
Qy	811	TTTAGCATCTCTTGGCTTGAGGACAGTGACCTTTCATGACATGAGCAAGAGACTTTGGAG	870
Db	830	TACAGCGTGTCTGGATCGAGATAGTGAGAGGCACAACTTGCACACGGAACTCTTTGAAG	889
Qy	871	CACAATATCACCGTTGTGTAAGAGAAGAGTAGGATCTGATGTACACAGAGACTTCTCATGT	930
Db	890	CAGCAATACCAAGTGGTTAAGAGCAAGACAGCAGAAAAAGCCTTCTACGCTCACATGTC	949
Qy	931	TGCGGTTTCGGACAGAGAGATGCTTAAAGATTATCTTTTCTCTTACATTTGAAGAAAT	990
Db	950	ATCGATGATGGTGATAAAGACTTACCGAGGACATGCTTTACTGTACATGGGTACAAAT	1009
Qy	991	CCTGAAAACGATAACTTCACCTTTCACGGAACTCTTTCTCCACCAATCTCTAATCTGGC	1050
Db	1010	CCTAATAATGAAAATTACACATATGTGGATGATAATTCGTTCGATCCTACATCTTCTAAT	1069
Qy	1051	TTGGTCAATCCGCGCGATATTCCTCTGCTATACCTCCAGAGAAGATTTCAAAAGCTCCA	1110
Db	1070	GCTGTTAATCAACGAGATGCCGATCTTATCCATTTCTGGAAACAAGTTTCGCAAGGCTTCT	1129
Qy	1111	ATGGGATCACTTGAAGCAAGAAGCTCAGAGAATAATGCTTCGACNAAGAATCATAGG	1170
Db	1130	GAAGGATCTCAAGAAGAAGATAAATGCCAGCAAGCAGTTTATGGAGGTGATGTCCACACA	1189
Qy	1171	AAACAAATCGATCAGAGCATTTACAGACATTTCTCGGCGTTTCAGTTAAACAACAACATGTC	1230
Db	1190	GTGCATTTGATGACAGCATCAAACTCATTTGGAAAGCTTCTATTTTGGAATCGAGAAGGC	1249
Qy	1231	TTAAATCTCTTAACCTTCCACAAGAACAAACAGGACAGCCTCTTGTAGACGATTTGGGATTG	1290
Db	1250	TTAGGTGTCTACAAACTGTCCGACCTACAGGACACACCCCTCGTCGATGACTGGAAGTGC	1309
Qy	1291	TTCAAGACTCTAGTTAATAGCTTCAAGAATCACTTCGGGTGCAACGGTGCATTACGGATTG	1350
Db	1310	CTGAAAACATTTGGTAAGGACATTTGAAAGCAATTTGTGGATCTCTATCTCAATACGGAATG	1369

Qy	1351	AAGTATACAGGAGCGCTTGC	AATATCTGCCAATATCGG	AGCTGGATGTGAAGCAAACTGTT	1410
Db	1370	AAACACATGCTTCGATTTG	CAACACATATGCGCGGGAT	CAACAATAATCAATATGGCC	1429
Qy	1411	TCAGCCATTGAACAAGCTT	GTTCGGA	1435	
Db	1430	GAGGCTTCAGCACAAAGCT	TGCCCGGA	1454	

Sequence

Sequence

RESULT	7
D89972	
LOCUS	D89972 1659 bp linear PLN 06-APR-1999
DEFINITION	Vigna mungo mRNA for asparaginyl endopeptidase (VmPE-1A), complete cds.

D89972
 D89972.1 GI:5489397
 asparaginyl endopeptidase (VMEP-1A).
 Vigna mungo seedling cotyledon cDNA to mRNA, clone:pPEB.
 Vigna mungo
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Vigna.

AUTHORS Okamoto, T. and Minamikawa, T.
TITLE Molecular cloning and characterization of Vigna mungo processing enzyme 1 (VnPE-1), an asparaginyl endopeptidase possibly involved in post-translational processing of a vacuolar cysteine endopeptidase (SH-EP)

JOURNAL	Plant Mol. Biol. 39 (1), 63-73 (1999)
MEDLINE	99178794
REFERENCE	2 (bases 1 to 1659)

TITLE
JOURNAL

Direct Submission
Submitted (19-DEC-1996) Takashi Okamoto, Tokyo Metropolitan
University, Department of Biology; Minami-osawa 1-1, Hachioji,
Tokyo 102-03, Japan (E-mail:okamoto-takashi@metro-u.ac.jp,
Tel:+81-426-77-2562, Fax:+81-426-77-2559)

FEATURES
source

```

source
1..1659
/organism="Vigna mungo"
/db_xref="taxon:3915"
/clone="pPEB"
/tissue_type="cotyledon"
/dev_stage="seedling"
1..45
5'UTR
/evidence=experimental
46..1494
CDS

```

5' UTR

CDS	/evidence=experimental
	46.1494

```

/ncbi_note="processing enzyme 1A"
/codon_start=1
/evidence=experimental
/product="asparaginyl endopeptidase (VnPE-1A)"
/protein_id="BAA76745.1"
/db_xref="GI:4589398"

```

```

/db_xref="GI:4589398"
/translation="MATTSLFLFLATVAARHDVDFLRSLPSGSDNDNV
QGTWRLPAGSYNMYRHOADI CHAYQILRKGLEENI IVFMDDIATFNWDRP
GVILNKPDGVDYGVYKDTGEDATAHNFSY SALLGDKSALTGSGKSVYSSGPDPRIF
IFYSDHGGCVGLCTPAGPIYASDLIVELVKKHASTYKNLVEYLACAGSTFEGLL
PBDINIYIATASNAESSWGTYCPGPPSPCYCLGLDLYSVAMWEDSDINLRKTE
SLHQQYKVKADRTQYGVGSHVYQGDVEFSKDALFLYLGTPDANDNLITFVDENSLW
SSATVNRDADLVHFMHFKRPAKSGPKKNEARKQOVLVMSHRMI IDSVKILVGLKL
FGFEKPEVLNVRPAGSLVLDWQACLKTMVRTFTHCGSLSQYGMKHMSPFANICNV
GIKKEQMAESAQAQCVTPASSWSNLQGFSA"
1495 165q
3'UTR

```

BASE COUNT	446 a	344 c	390 g	479 t
ORIGIN	/evidence=experimental			

Query Match 29.9%; Score 466; DB 8; Length 1659;
Best Local Similarity 60.8%; Pred. NO. 2.4e-95;
Matches 779; Conservative 0; Mismatches 500; Indels

Matches	779	Conservative	0	Mismatches	500	Indels	3	Gaps	1
<p> </p>									


```

/Note="This clone is in puni 51
ecotype: Columbia"
1. 1485
/Note="gamma-VPE (vacuolar processing enzyme)"
/codon_start=1
/product="At4g32940/F26P21_60"
/protein_id="F26P21_60"
/db_xref="GI:22137032"
/translation="MATTMTKRSVGVVLFVLVLVLVAARSAPDVBKIPQASRF
FRPADDNSNGTFRNVLVAGSGGYWNTYRHQADICHAIVQLLRGGKLEENIVVFMYD
DIANNENPRPGTIIINSRPGKDYOCVQKDYDNDVNDVNFVILGDKTAVGSGSK
WDSGPNDRHIFYSRHPGKDYOCVQKDYDNDVNDVNFVILGDKTAVGSGSK
CDS
365 c 365 g 403 t
BASE COUNT
ORIGIN
29 28; Score 454.8; DB 8; Length 1485;
Best Local Similarity 60.58; Pred. No. 8.3e-93;
Matches 785; Conservative 0; Mismatches 507; Indels 6; Gaps 2;
QY 141 TGTGAATCTAGCAGCAAGAGTGCAGAAAGCCACAGATGGGCTGTTTGTGCTGATC 200
DB 141 TGTGAATCTAGCAGCAAGAGTGCAGAAAGCCACAGATGGGCTGTTTGTGCTGATC 200
QY 201 AATGAATATATTAAGTACAGGATCAGGCTGACATATGCCACGGTATCAGATCTCCG 260
DB 201 AATGAATATATTAAGTACAGGATCAGGCTGACATATGCCACGGTATCAGATCTCCG 260
QY 261 AAAGCGGCTTTAAAGATGAAACATCATTTGTTGTTTATGATGATATATCGCGTTTTC 320
DB 261 AAAGCGGCTTTAAAGATGAAACATCATTTGTTGTTTATGATGATATATCGCGTTTTC 320
QY 321 CTCGAGAGATCTAGGCTCGAGTTCATTTAATAAACAGATGAGAGATGTTTATAA 380
DB 321 CTCGAGAGATCTAGGCTCGAGTTCATTTAATAAACAGATGAGAGATGTTTATAA 380
QY 381 AGAGTCTTAAAGATGAAACATCATTTGTTGTTTATGATGATATATCGCGTTTTC 440
DB 381 AGAGTCTTAAAGATGAAACATCATTTGTTGTTTATGATGATATATCGCGTTTTC 440
QY 441 ACTTGAATGAAAGTGGCTACAGGAGGAATGCGAAGTTGGAAGTGGTCTCTAA 500
DB 441 ACTTGAATGAAAGTGGCTACAGGAGGAATGCGAAGTTGGAAGTGGTCTCTAA 500
QY 501 TGATATATCTTATATATGCTGACCATGAGCTCTGCTCTTAATAGCGATGCCCCAC 560
DB 501 TGATATATCTTATATATGCTGACCATGAGCTCTGCTCTTAATAGCGATGCCCCAC 560
QY 561 TGTGATGAAGTTATGCGAAAGATTTCAATGAAGTCTTGGAGAGATGCATAAGAGAAA 620
DB 561 TGTGATGAAGTTATGCGAAAGATTTCAATGAAGTCTTGGAGAGATGCATAAGAGAAA 620
QY 621 AAATACAAAGATGTTGATCTATGTTGAAGCATGTGAATCAGGAGTATGTTGAAG 680
DB 621 AAATACAAAGATGTTGATCTATGTTGAAGCATGTGAATCAGGAGTATGTTGAAG 680
QY 681 GATTTAAGAAATATCAACATATACCGAGTCTGCTGATATTTCAAGAGACAG 740
DB 681 GATTTAAGAAATATCAACATATACCGAGTCTGCTGATATTTCAAGAGACAG 740
QY 741 CTGGGAGTTTACTGCTGAGTCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 800
DB 741 CTGGGAGTTTACTGCTGAGTCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 800
QY 801 CGGCGATACATTTAGCATCTCTGCTGTTGAGGACAGTACCTTTCATGATGACAGCAAGA 860
DB 801 CGGCGATACATTTAGCATCTCTGCTGTTGAGGACAGTACCTTTCATGATGACAGCAAGA 860
QY 861 GACTTTGGAGCAACATACCAAGCTTGTAAAGAGAGAGTAGGATCTGATGATC --- CAGA 917

```

```

861 GACTCTGCACCAAGATATGAACCTTGTGAAAGGAGGACTGCACCTGTTGGTACTCTTA 920
QY 918 GACTCTGCATGATGATCCGCTTTTCGGAAGAGAGAGATGCTTTAAAGATATATCTTCTCTTA 977
DB 921 TGGTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 980
QY 978 CATTTGAAGAAATCTTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1037
DB 981 TATTTGAAGAAATCTTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1037
QY 1038 CTCTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1097
DB 1038 GCAACCTTCTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1097
QY 1098 TCAAAAGCTTCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1157
DB 1098 CCGAAAGCTTCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1157
QY 1158 AAGAAATCATAGAAACAAATCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1217
DB 1158 CATGCTCAGACATCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1217
QY 1218 ACAACCAATGCTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCT 1277
DB 1218 CATTTGAGAGGCTTGAAGTCTTGAAGTCTTGAAGTCTTGAAGTCTTGAAGTCTTGAAGTCT 1277
QY 1278 CGATTGGGATCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCT 1337
DB 1278 TGACTGGAATCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCT 1337
QY 1338 GCATTACGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1397
DB 1338 TCAGTACGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCT 1397
QY 1398 GAAGCAATCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCT 1435
DB 1398 GGAGCAATCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCAAGAGCT 1435

RESULT 10
AY059104 1516 bp mRNA linear PLN 18-APR-2002
LOCUS Arabidopsis thaliana putative vacuolar processing enzyme gamma-VPE
DEFINITION (At4g32940) mRNA, complete cds.
ACCESSION AY059104
VERSION AY059104.1 GI:16323431
KEYWORDS FLI_CDNA.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1516)
AUTHORS Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Elgu, P.,
Lee, J.M., Toriumi, M., Yu, G., Brooks, S., Chao, Q., Chen, H.,
Karlin-Neumann, G., Kim, C., Lam, B., Miranda, M., Nguyen, M.,
Palm, C.J., Shinn, P., Southwick, A., Davis, R.W., Ecker, J.R. and
Theologis, A.
TITLE Arabidopsis Open Reading Frame (ORF) Clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1516)
AUTHORS Yamada, K., Banh, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D.,
Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C.,
Yanamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H.,
Chen, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A.,
Karlin-Neumann, G., Kawai, J., Kim, C., Koesema, E., Lam, B., Lin, J.,
Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,
Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S.E.,
Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA

```

COMMENT	The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.	
	The Salk, Stanford, PGECC (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAFL cDNAs: Yamada, K., Banh, J., Banno, F., Chang, E., Dale, J. M., Goldsmith, A. D., Lee, J. M., Onodera, C. S., Quach, H. L., Tang, C. C., Toriumi, M., Wu, H. C., Yamamura, Y., Yu, G. S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M. C., Miranda, M., Nguyen, M., Palm, C. J., Shinn, P., Southwick, A., Tracy, S. E., Davis, R. W., Ecker, J. R. and Theologis, A.	
FEATURES	Yamada, K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGECC) contributed equally to this work as PIs.	
	Location/Qualifiers	
source	1. .1516	
	/organism="Arabidopsis thaliana"	
gene	/db_xref="taxon:3702"	
	/chromosome="4"	
CDS	/clone="U10153"	
	/note="this clone is in pUNI 51."	
3'UTR	ecotype: Columbia"	
	1. .1516	
BASE COUNT	/gene="At4g32940"	
	1. .1485	
ORIGIN	/gene="At4g32940"	
	/codon_start=1	
Query Match	/evidence=experimental	
	/product="putative vacuolar processing enzyme gamma-VPE"	
Best Local Similarity	/protein_id="AA15210.1"	
	/db_xref="GI:16323432"	
Matches 785; Conservative 0; Mismatches 507; Indels 6; Gaps 2;	/translation="MATTMTSVGVVFLVLLVSLVAVSAARSPPDDVIKLSQASRF FRANDDSSNCTRAVLVAGSGWYVHROADICHAYOLLKRGGLKENIVFYFD DIANNENPRFIIYSOHGGPVLGMPTSPYLIANDLNLKRLHAGTYSKLVFLYEA VDSGPNNDHIFISDHGGPVLGMPTSPYLIANDLNLKRLHAGTYSKLVFLYEA CESGIFEGLLPGLIYATTNAEBSWGTCPGPEEPSPPPEYETCLGLDLYSVAM EDSGMHLQETLHQOYELVKRTPAVGYSYGVHVMQYGDVPSKNDLDMYMTNPAN DNFTFADANSLPQVTVNORDADLVHFEKRYKAPGSKARKTEAKOKVLEAMSHRLH IDNSVILVGLFISRGPELVNKRVSAGLOPLVDDNCLKNQVRAFERHCGSLSQYGI KMRSFANICNAGIOMQEAEASQACTTTLTPGWSSLNKRFSA"	
	1486. .1516	
418 a 316 c 368 g 414 t	/gene="At4g32940"	
Qy	29.28; Score 454.8; DB 8; Length 1516;	
	60.58; Pred. No. 8.3e-93;	
Db	141 TGTGTAATCTAGGCACAGAGTGCACACAGTGGCTGTTTCTAGTGTGGATC 200	
	141 TCGTGAACACGACGATCTTAACTCCGCTACTAGTGGCTGTTCTAGTCCCGGATC 200	
Qy	201 AAATGAATATATTAATCTACAGGATCAGGCTGACATATGCCAGCGTATCAGATACCTCG 260	
	201 TAGCGGATATGGAAATACAGGATCAGGCTGATATATGCCATCCCTATCAACTTCTGAG 260	
Qy	261 AAAGGCGGTTTAAAGATCAAAACATCATGTGTTTATGTATGATGATATCGGTTTC 320	
	261 GAAAGGTGGATTGAAGAGGAGATATTTGGTATTTCATGATGATATTCCTAACAA 320	
Qy	321 CTGCGAATCTTAGGCTGGAGTTATCATTAATAAACACAGATGGAGAACATGTTTATAA 380	
	321 TTACGAGAATCCAAAGGCTCGAACCATATACACAGCCCTCATCGGAAGATGCTATCA 380	
Qy	381 AGAGTTCTTACAGGATCTACTAAAGAGCTGTGAATGTTTCAAACTTCTACAACTGTGT 440	
	381 AGGAGTCCCAAGGATATATCGAGATGATGTCATGTTGATAATCTATTCTGCTGTGAT 440	

RESULT 11
AF424619
LOCUS
DEFINITION

Arabidopsis thaliana AT4g32940/P2621_50 mRNA, complete cds.

1760 bp mRNA linear PLN 08-OCT-2001

Best Local Similarity 60.5%; Pred. No. 4.8e-92; Matches 778; Conservative 0; Mismatches 501; Indels 6; Gaps 2;									
QY	150	TAGCGACAAGAGTGCRAAAGGCGACAGCATGGCTGTTTATAGTTGCTGGGATCAAAATGAATA	209	1163	CAGAATGCATATAGACGACAGTGTGGAACTTTGTTGAAAGACTTTTATTTGGCATTTGAAA	1222	1227	TGCTTTAAATCTCTTAACCTTCCACAAGAACAAGACGACGCTCTTTGTAGACGATTTGGGA	1286
DB	146	TAACGGGACAACGTCATCGAAGACGCGCTGGGCCATCTCTTCGCTGGTTCGAGCGGCTA	205	1223	GGCTCTCTGAATTAATGCTGTAGACCGGCTGGATCCGCACTTGTGTGATGACTGGGA	1282	1287	TTGCTTCAAGACTCTAGTTAATAGCTTCAAGAATCACTCGGTCGCAACGGTGCATTTACGG	1346
QY	210	TTATAACTACAGGCATCAGGCTGACATATATGCCGCGTATFAGATCACTCCGAAAGCGGG	269	1283	CTGCTTGAAGCAACCATGTAAGGACTTTTGGAGACACATTTGTGGATCTGTCTCTCAGTATGG	1342	1283	CTGCTTGAAGCAACCATGTAAGGACTTTTGGAGACACATTTGTGGATCTGTCTCTCAGTATGG	1342
DB	206	TTGGAATATACAGACATCAGGCCGATATTTGCTATGCTATCAATTAATCTAGGAAAGGTGG	265	1347	ATTGAAGTATACAGGAGCGCTTGCCAATATCTCAATATGGGAGTGGATGTGAAGCAAC	1406	1347	ATTGAAGTATACAGGAGCGCTTGCCAATATCTCAATATGGGAGTGGATGTGAAGCAAC	1406
QY	270	TTTAAAGATGAACATCATCTGTTTATGTATGATGATATCGCGTTTTCCTCGGAGAA	329	1343	GATGAACACATGAGGTCTTTTGGACATGTGCAATGTAGGAATAAAGAAGGAGCAAT	1402	1343	GATGAACACATGAGGTCTTTTGGACATGTGCAATGTAGGAATAAAGAAGGAGCAAT	1402
DB	266	CCTGAAGATGAATAATCATCTTTTATGTATGATGACATTCATTTTCGGA	325	1407	TGTTTCAGCCATTGAACAAGCTTGT	1431	1407	TGTTTCAGCCATTGAACAAGCTTGT	1431
QY	330	TCCTAGGCTTGGAGTATCATTAATAAACCAGATGGAGAAGATGTTTATAAAGGAGTTC	389	1403	CGGTGAAGCCTCAGCACAAAGCTTGT	1427	1403	CGGTGAAGCCTCAGCACAAAGCTTGT	1427
DB	326	TCCAAGCGTGGAGTCATCATTAACAGTCCAAATGGAGATGAAGTTTATAAAGGAGTTC	385	RESULT 14					
QY	390	TAAGGACTACACTAAAGAAGCTGTTAATGTTCAAAACCTCTACAATGTGTACTTGGAA	449	E05717	LOCUS				
DB	386	AAAGGATTAACCTGGTGAAGATCTTACTGCTCATAACTTTTATGCTGCTTACTTGAGA	445	DEFINITION	Asparaginilendopeptidase gene.				
QY	450	TGAAAGTGGCTCACAGGAGGAATGGCAAAGTTGTCAAAGTGGTCCCTAATCATATAT	509	ACCESSION	E05717				
DB	446	TAAGTGAAGTACTGGTGGCAGTGGGAAAGTTGTGAATAGTGGGCCCAATGATCACAT	505	VERSION	E05717.1 GI:2173904				
QY	510	CTTCACTATTATGCTGACCATGGAGCTCTGGCTTAATAGCGATGCCCACTGGTGATGA	569	KEYWORDS	JP 1993276960-A/2.				
DB	506	ATTATATTACTCTGATCATCTGCTGCTCAGGGTCTCGGGTCCGCTGCTGCTCTTA	565	SOURCE	Canavalia ensiformis.				
QY	570	AGTTATGGCAAGAAGTTTCAATGAAGTCTTTGGAGAAGATGCATPAAGAGAAAATAACAA	629	ORGANISM	Canavalia ensiformis				
DB	566	CATTTATGCATCTGCTGCTGAATGAAGTATTGAAGAAAACATGCTTCTGGAACGTATA	625	REFERENCE	1 (bases 1 to 1323)				
QY	630	CAAGATGGTATCATGTTGAACATGTGAATCAGGAAGTATGTTTGAAGGGATTTTAA	689	AUTHORS	Mita,M., Takeda,O., Katou,I., Ishii,S. and Abe,Y.				
DB	626	AAACCTAGTATTTTATTTGGAGCATGTGAATCTGGAAGTATCTTTGAAGGTCTTCTTC	685	TITLE	ASPARAGINYL ENDOPEPTIDASE GENE				
QY	690	GAAAAATCTCAACATATACGACGTGCTGCTTAATTTCAAAGACAGCAGCTGGGAGT	749	JOURNAL	Patent: JP 1993276960-A 2 26-OCT-1993;				
DB	686	TGAAGATATCAATGTTTATSCAACACAGCTTCAATGCGGATGAAGTATTTGGGAAC	745	COMMENT	TAKARA SHUZO CO LTD				
QY	750	TTACTGCTGATCATATCTCTCTCTCTGAGATTTGAAGTGTGCTCGGCGATAC	809	OS	Canavalia ensiformis				
DB	746	ATATTGCTCTGGGAGGATCTAGTCTCTCCCAAGAAATATTCACCTGCTTGGGTGACCT	805	PN	JP 1993276960-A/2				
QY	810	ATTTAGCATCTCTGGCTTGAGACAGTGAACCTTTCATGACATGAGCAAGAGACTTTGA	869	PD	26-OCT-1993				
DB	806	GTACAGTGTGCTTGGATGGAACACAGTGTAGACAAATTTGCGAACAGAACTTTGCA	865	PF	07-AUG-1992 JP 1992231602				
QY	870	GCAACAATACACGTTTGAAGAGAGAGTGGATCTGTATACCAGAGACT---TCTCA	926	PR	07-FEB-1992 JP 92P 56023				
DB	866	CCAACAATAAATTTGTTAAAGAGAGACTATATCTGGAGGTTTATCTATGCTCTCA	925	PI	MITA MASANORI, TAKEDA OSAMU, KATOU IKUNOSHIN, ISHII SHINICHI.				
QY	927	TGATGCGGTTTCGGAACAGAGAGATGCTTAAAGATTAATCTTTCCTCTTACATGGAAG	986	PI	ABE YUKICHI				
DB	926	CGTATGTCAGTATGGTATGATGGGCTTAGCAAGGATATCTCTTCCATTTATTTGGGTAC	985	PC	C12N15/57,C12N9/48;				
QY	987	AAATCTGAAAGCATTAACCTTCACTTTACGGGATCTTTTCTCCTACCAATCTCTAATTC	1046	CC	strandedness: Double;				
DB	986	AGATCTGCAATAGGAATTTGACTTTCTGGGATGAAAGTCTCT---TATGGTCACTCTC	1042	CC	topology: Linear.				
QY	1047	TGCTTTGGTCAATCCGCGCATATCTCTGCTATACCTCCAGAGAAAGATTCAAAAGC	1106	FEATURES	Location/Qualifiers				
DB	1043	AAAGCAGTCAACCAAGGATCTGATCTCTGCTCAATTTTGGGATGAAGTTCGCAAGC	1102	source	I..1323				
QY	1107	TCCAATGGGATCACTTGAAGCAAGAGCTGAGAGAAATGCTTTGACGAAAGAAATCA	1166	Query Match	28.5%; Score 444; DB 6; Length 1323;				
DB	1103	TCCTGAGGGTCTCTCTTAAGAAAAATGAAGTCGGAAGAGAGTCTCGAAGTAAATGCTCA	1162	Best Local Similarity	60.5%; Pred. No. 2.4e-90;				
QY	1167	TAGGAACAAATGATCAGAGCATTTACAGACATTTCTGGGCTTTTCAGTTAAACAAACCA	1226	Matches	767; Conservative 0; Mismatches 495; Indels 6; Gaps 2;				

DHGGPVLGMPSTPNLYANDLNDLVKKYASGYKSLVFLYLEACSSIFEGLLPEGL
NIYATTASNAEPSSGYCGEDSPPESEYETCLGLDLYSVAMIEDSKHNLQETLHE
QYELVKRTRTAGSKSGYSGHVMFEDIGLSKEKLVFMGTNPADENFTFVENSIRPPS
RVTRORDADLVHFWHKYQKAPEGSARKVEAKOVLEAMSHRLHVDNSILLIGILLFGL
EGHAVLNKVRPSGEPLVDNDCLSLVRAFERHCGSLSQYGIKHKRSIANNWCNAGIQM
ROMEAMQACPTPIPTSPWSLDRGFS*

BASE COUNT 400 a 264 c 361 g 412 t
ORIGIN

Query Match 28.3%; Score 442.2; DB 8; Length 1437;
Best Local Similarity 59.6%; Pred. No. 6e-90;
Matches 801; Conservative 0; Mismatches 533; Indels 9; Gaps 3;

QY	96	CTTCTCAGCTGAGTCCCGCAAAACC	CAATTCCTGCAACGATATGATGTGAATCTAGCGA	155
DB	51	CGTTTCTGGTCACGTCATCAAACT	TCCTCTCTAGCTTCTTAAGTCTTCCTCCGCCCAACTGA	110
QY	156	CAAGAGTGCAAAAGGCACACAGAT	GGCGTGTCTTGTAGTTCGTGGATCAAAATGATATTATAA	215
DB	111	AAACGACGATGATTCTACTAAGT	GGCGTGTCTTCGTCCGCGGATCCAGCGGATACTGGAA	170
QY	216	CTACAGGCATCAGCTGCATATG	CCAGCGGTATCAGATATCTCCGAAAAGCGGTTTAAA	275
DB	171	TTATCGTCATCAGCGGATGTTGT	CTCATCTTATCAGCTTTTGAAGAAGGTGGAGTGAA	230
QY	276	AGATGAAACATCATTTGTGTTAT	GTATGATGATATCCGCGTTCCTCGGGAATCTCTAG	335
DB	231	AGAGGAGAATATCTGGTGTGTT	TATGATGATGACATTCGGAAGAAGCAGGAGAGATCCAAG	290
QY	336	CCCTGGAGTTATCATTAATAAC	CAGATGGAGAGATGTTTATAAGAGGTTCTTAAGGA	395
DB	291	ACCTGGAGTTATATCATATG	CTTAATGGAGAGATGCTCTATGAGGATGCCCAAGGA	350
QY	396	CTACACTAAAGAAGCTGTTAAT	GTTCAAAACCTCTACAAATGTGTACTTTGGAATGAAG	455
DB	351	TTACACTGGAGATGAAGTTAAT	GTGTGAACCTTATAGCTGTGATCTTGGAAACAAAC	410
QY	456	TGGCGTCACAGGAGAAATGCA	AAAGTTGTGAAAGTGGTCCCTAATGATAATATCTTCAT	515
DB	411	GGCTCTTAAAGGAGAGAGTGG	AAAGTTGTAGATAGCGGTCCAAACGATCATATCTTTAT	470
QY	516	CTATTATGCTGACCATGGAG	CTCTCGGCTTAATAGCGATGCCACTGGTGATGAAGTTAT	575
DB	471	ATAGTATAGTATCACGGTGT	CCGGAGTCTCGGATGCCAACTCTCCAAAACCTATA	530
QY	576	GGCAAAAGATTTCAATGAAG	CTTGGAGAAGATGCATAAAGAAAAAATACAAACAGAT	635
DB	531	TGCAAAATGATCTCAATGAT	GTCTTGAAGAAAAATATGCTTCAGGAACATATAAGAGCTT	590
QY	636	GGTGATCTATCTTGAAGCAT	GTGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAA	695
DB	591	GGTGTGTTTATTTGGAGG	CTTCTGAATCTGGAAGTATTTTTCGAAGTCTTTTACCAGAGG	650
QY	696	TCTCAACATATACGAGTACT	CGTCTGCTAATTTCTAAGAGAGCAGCTGGGGAGTTTACTG	755
DB	651	TTTAAATATTTACCGCAAC	ATGTCATGAAATGAGAAAAAGTAGCTGGGGTACTTACTG	710
QY	756	TCCTGAGTCATATCTCTCT	CTCTCTGAGATTTGGAACCTGTCTCGGCGATACATTTAG	815
DB	711	TCCTGGAGAGGATCCTAG	TCTCTCTCTGAGTATGAGACCTGTGTTGGTGACTTATACAG	770
QY	816	CATCTCTTGGCTTGAGG	ACTGACCTTTCATGACATGAGCAAGAGACTTTGGAGCAACA	875
DB	771	TGTTGCTTGGATAGAAG	ACTAGTAAACAACTTTTACAAAACAGAGACTTTGGCAGGACA	830
QY	876	ATACCACCTTGT ---	AAGAGAAAGTAGGATCTGATGTACACAGACTTCTCATGTATG	932
DB	831	ATATGAATTTGGTGA	AAAAAGAGAACTGCAGGTTCTGGTAAGTCTTATGGTCTCATGTAT	890
QY	933	CCGTTTCGGAACAGAG	AGATGCTTAAAGATTTATCTTCTCTTACATTTGGAAGAAATCC	992
DB	891	GGAAATTTGGAGATAT	AGGACTCAGCAAGGAGAGCTTGCTCTCTTTTATGGGTACAAATCC	950

Search completed: May 29, 2003, 16:32:08
Job time : 4045 secs

QY	993	TGAAAACGATAACTTCACTTT	CACGGAATCCCTTTCTCACC	AAATCTCTTAATTTCTGGCTT	1052
DB	951	AGCAGATGAAAACTTACCT	TTTGTGAATGAGAAAT---T	CAATAAGGGCGCCTTCAAG	1007
QY	1053	GCTCAATCCGCGCATATTC	CTCTGCTATACCTCCAGAGA	AAAGATTTCAAAAAGCTCCA	1112
DB	1008	TACAAACGAGGTGATCGG	ATCTTGTCCTTCTGGCAT	AAAGTATCAAAAAGGCACCG	1067
QY	1113	GGGATCACTTGAAAGCA	AAAGAGCTCAGAAAGAAAT	TGCTTCAGCAAAAGAAATCA	1172
DB	1068	AGGTCAGCAAGAAAAGT	TGAAGCTCAGAACAAAGT	CTTGAAGCAATGTCTCAC	1127
QY	1173	ACAAATCGATCAGAGCAT	TACAGACATTCCTCGGG	CTTTTCAGTTAAACAAACCA	1232
DB	1128	TCATGTTGATAATAGCA	TTCTGTTGATGGGATCT	TTTGTGTTT---GGTTTGA	1184
QY	1233	AAATCTCTTAACCTTCC	ACAGAACAGGACAGCC	CTCTTCTAGACGATTTGGG	1292
DB	1185	TGCGGTGTAAATAAG	TCGCGCTTCTGGAGAAC	CCGCTTGTGACGATTTGGG	1244
QY	1293	CAAGACTCTAGTTAAT	AGCTTCAAGAAATCACT	CGGCTGCAACGGTGCAATT	1352
DB	1245	TAAATCTCTGTGAG	AGCTTTTCGAGAGC	ACTGTGGATCGTTGTCT	1304
QY	1353	GTATACAGGAGCGGT	TGCAATATCTGCAAT	TATGGGAGTGGATGTGA	1412
DB	1305	GCACATGAGGTGAT	TGCAAAACATGTG	CAACGAGGATTCAGAT	1364
QY	1413	AGCCATTCAACAG	CTTCTTTCGA	1435	
DB	1365	GGCAGCAATG	CAGGCTTGTC	1387	

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: May 29, 2003, 17:10:14 ; Search time 3075 Seconds
(without alignments)
4410.376 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFOILVFLHLLIF.....CNMGVDVKQTVSAIEQACSM 466

Scoring table: BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US0934066/runat_19052003_163659_24353/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0934066 @cgn_1_1_3745 @runat_19052003_163659_24353 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: gb_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_mu.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

Result No.	Score	Match %	Query Length	DB ID	Description
1	2454	100.0	1563	8	AY120765 Arabidops
2	2446	99.7	1541	8	AY084227 Arabidops
3	2188	89.2	83253	8	AP000383 Arabidops
4	1269.5	51.7	1724	8	D89971 Vigna mungo
5	1268.5	51.7	1834	8	VSCYSPROB
6	1262.5	51.4	1839	8	CSCYSPRNS
7	1254.5	51.1	1576	8	AF238384
8	1252	51.0	1659	8	AF238384
9	1246	50.8	1485	8	D89972
10	1246	50.8	1516	8	AY133531 Arabidops
11	1246	50.8	1760	8	AF059104 Arabidops
12	1246	50.8	1779	8	AF24619 Arabidops
13	1243.5	50.7	1850	8	AF370160 Arabidops
14	1236	50.4	1980	8	PV299956
15	1230.5	50.1	1437	8	BV0309173
16	1230.5	50.1	1616	8	AY090296 Arabidops
17	1228.5	50.1	1866	8	AY058055 Arabidops
18	1220	49.7	1323	6	AB025310 Oryza sat
19	1219	49.7	1572	8	E05717 Asparaginil
20	1215	49.5	1323	6	AF260827 ipomoea b
21	1208.5	49.2	1919	8	E05718 Asparaginil
22	1191.5	48.6	1864	8	AF082347 Zea mays
23	1183.5	48.2	1569	8	AJ131719 Zea mays
24	1166.5	47.5	1529	8	AJ131718 Zea mays
25	1160	47.3	1461	8	LES243876 Lycopersi
26	1160	47.3	1643	8	AY059156 Arabidops
27	1118.5	45.6	1649	8	AF367254 Arabidops
28	1115.5	45.5	1652	8	AF169973 Sesamum i
29	1109	45.2	1762	8	PV299957
30	1094	44.6	1140	8	RCCVPE
31	1092.5	44.5	1642	8	VNA238882
32	1080	44.0	1323	6	AY062178 Oryza sat
33	1080	44.0	1323	6	E05716 Asparaginil
34	1080	44.0	1836	8	E05720 Asparaginil
35	1080	44.0	1840	8	D31787 Jack Bean m
36	1072.5	43.7	1437	8	VNCYSTPRO
37	1071	43.6	1736	8	NTA238880
38	1054	43.0	1851	8	SOYCPA
39	1051.5	42.8	1360	8	VSA007743
40	1051	42.8	1485	8	NTA238881
41	1046.5	42.6	1715	8	AF082346
42	1035	42.2	45160	8	AF169019 Glycine m
43	1035	42.2	143379	2	AC130727 Oryza sat
44	1029	41.9	656	8	AC130728 Oryza sat
45	977	39.8	1152	6	AY070375 Arabidops
					E05721 Asparaginil

ALIGNMENTS

RESULT 1

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

AY120765 1563 bp mRNA linear PLN 22-JUN-2002
 LOCUS Arabidopsis thaliana vacuolar processing enzyme/asparaginyl
 DEFINITION endopeptidase, putative (At3g20210) mRNA, complete cds.
 ACCESSION AY120765
 VERSION 1
 KEYWORDS
 SOURCE FLI-CDNA.
 ORGANISM Arabidopsis thaliana
 thale cress.
 REFERENCE 1 (bases 1 to 1563)
 AUTHORS Southwick,A., Karlin-Neumann,G., Nguyen,M., Tripp,M., Miranda,M.,
 Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
 Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
 Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
 Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
 TITLE Direct Submission
 JOURNAL Submitted (11-JUN-2002) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 COMMENT e-mail for correspondence: arab@sequence.stanford.edu
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA: "RIKEN
 Arabidopsis Full-length cDNA"): Seki,M., Narusaka,M., Ishida,J.,
 Hayashizaki,Y. and Shinozaki,K.
 The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Southwick,A.,
 Nguyen,M., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J.,
 Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
 Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P.,
 Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.
 Southwick,A. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
 equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W.
 (SSP/Stanford) contributed equally to this work as PIs.
 FEATURES
 source
 1. 1563
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="RAFL09-06-A19 (R16622)"
 /note="This clone is in pBluescript
 ecotype: Columbia"
 gene
 1. 1563
 /gene="At3g20210"
 /note="MAL21.23"
 CDS
 63..1463
 /gene="At3g20210"
 /codon_start=1
 /product="vacuolar processing enzyme/asparaginyl
 endopeptidase, putative"
 /protein_id="AA053323.1"
 /db_xref="GI:21539541"
 /translation="MSSPLGHQILVFLHALLIFSAESRKTQLLNDVSSDKSAG
 TRWVLAAGSNEYNYRHQADICAYQILRKGGLKDENIIVFMDIAFSGNPFCV
 LNKPDGDVYKPKQDYKQVAVQNFYVLLNGESVGTGKGVKVGSGNPDIYI
 YADKAPGLIAMPQDVEYMAKDFNEVLEKMKRKYKMKVIYVEACESGSMFEGILK
 NLNIVYVAANSKESMGVYCPSPPPPEIGTCLDGTFSISWLEDSDLHDMSKETL
 EQQHVAVRVSQVPETSHVCRFGTEKMLKDYLSYIGNPENDNFTFESESPIS
 NSGLNPRDIPLLYLRKIQAKPMGSLKESKQAOKLLDEKNRKOIQDQITDILRSV
 KOTNVLNLTSTRTTGGPFLVDDMDCFKTLVNSFKNHCATVHYGLKTYGTALANICNG
 VDVQTVSAIEQAQCSM"
 BASE COUNT 508 a 290 c 322 g 443 t
 ORIGIN
 Alignment Scores: 3.02e-208 Length: 1563
 Pred. No.: 1

Score: 2454.00 Matches: 466
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0
 US-09-934-066-2 (1-466) x AY120765 (1-1563)
 QY 1 MetSerSerProLeuGlyHisPheGlnLeuValPheLeuHisAlaLeuLeuLeuPhe 20
 Db 63 ATGTCTAGTCTCTTGGTGCACTTTTCAGATTCTTGTGTTTCTTCATGCTTTCATCTTC 122
 QY 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAspLys 40
 Db 123 TCAGCTGAGTCCCGCAAAACCAATTCCTGAACGATAATGATGTTGAATCTACGCAAG 182
 QY 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
 Db 183 AGTCAAAAGGCACACGATGGCTGTTTATGTTGCTGATCAATATTAATACTAC 242
 QY 61 ArgHisGlnAlaAspPheCysHisAlaTyrClnIleLeuArgLysGlyGlyLeuLysAsp 80
 Db 243 AGGCATCAGGCTGACATATGCCACGCTATCAGATACCTCCGAAAGCGGCTTTAAAGAT 302
 QY 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
 Db 303 GAAACATCATTTGTTTATGTATGATGATATCCGTTTTCCTCGGAGATCTCTAGCCCT 362
 QY 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
 Db 363 GGAGTTATCATTAATAACACAGATGGAGAGATGTTTATAAGAGAGTTCCTAAGGACTAC 422
 QY 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
 Db 423 ACTAAGAAGCTGTTAATGTTCAAAACCTCTTACAAATGTTTACTTGGAAATGAAGTGC 482
 QY 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
 Db 483 GTCACAGAGGAAATGGCAAAAGTTGTGAAAAGTGGTCTTAATGATATATCTTTCATCTAT 542
 QY 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
 Db 543 TATGCTGACCATGGAGCTCCTGGCTTAATAGCGATGCCACTGGTGATGAAGTTATGGCA 602
 QY 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200
 Db 603 AAAGATTTCAATGAAGTCTTGGACAACATGCATAAGAGAGAAAAATACACAAGATGGTG 662
 QY 201 IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
 Db 663 ATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAAATCTC 722
 QY 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPro 240
 Db 723 AACATATACCGAGTCACTGCTGCTAAATTCATAAGAGAGACGCTGGGAGGTTTACTGTCT 782
 QY 241 GluSerTyrProProProProSerGluIleGlyPheCysLeuGlyAspThrPheSerIle 260
 Db 783 GAGTCATATCT 842
 QY 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
 Db 843 TCTTGGCTTGAAGACAGTACGCTTTCATGACATGACGAAAGAGACTTTGGAGCAACAATAC 902
 QY 281 HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
 Db 903 CACGTTGTAAGAGAGAGAGTATGATCTGATACACAGAGACTTCTCATGTATGCCCTTC 962
 QY 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
 Db 963 CGAACAGAGAGAGTCTTAAAGATTATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1022
 QY 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuLeuAsn 340

```

Db      1023 GATAACTTCACTTTCACGGAATCCTTTTCTCACCATCTCTAATTCGGCTTGGTCAAT 1082
Qy      341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db      1083 CCGCGGATATTCCTCTGCTATACCTCCAGAGAAAGATTCAAAAGCTCCAAATGGGATCA 1142
Qy      361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db      1143 CTTGAAGACAAAGAGCTCAGAGAAATTCCTTGACGAAAGAAATCATAGAAACAATC 1202
Qy      381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db      1203 GATCAGAGCATTCAGACATCTTCGGCGCTTTCAGTTAAACAAACCAATGCTTAAATCTC 1262
Qy      401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db      1263 TTAACCTCCACAGAACACAGACAGCCCTCTTGAGACGATGGGATGCTTCAAGACT 1322
Qy      421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db      1323 CTAGTTAATAGCTCAAGAAATCACTGCGGTGCAACGGTGCAATACGGATTGAAGTATACA 1382
Qy      441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db      1383 GGAGCGCTTCCCAATATCTGCAATATGGGATGGATGTGAAGCAAACTGTTTCAGCCATT 1442
Qy      461 GluGlnAlaCysSerMet 466
Db      1443 GAACAAGCTTGTTCGATG 1460

RESULT 2
LOCUS      AY084227                1541 bp. mRNA linear PLN 21-JUN-2002
DEFINITION Arabidopsis thaliana clone 100372 mRNA, complete sequence.
ACCESSION  AY084227
VERSION     AY084227.1 GI:21402937
KEYWORDS   FLI_CDNA.
SOURCE     Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
            1 (bases 1 to 1541)
            Haas,B.J., Volkovsky,N., Town,C.D., Troughan,M., Alexandrov,N.,
            Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
            Full-length messenger RNA sequences greatly improve genome
            annotation
JOURNAL    Genome Biol. (2002) In press
REFERENCE  2 (bases 1 to 1541)
            Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
            Full-length cDNA from Arabidopsis thaliana
JOURNAL    Unpublished
REFERENCE  3 (bases 1 to 1541)
            Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
            Direct Submission
JOURNAL    Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
            Malibu, CA 90265, USA
COMMENT    This clone sequence is one of 5,000 Ceres full-length cDNAs made
            available to TIGR and Genbank. The following quality assessment of
            this set was done by comparison with known proteins: two percent of
            the clones are estimated to be 5'-truncated; less than one percent
            are 3'-truncated; approximately two percent represent alternative
            splice variants, including unspliced introns and spliced exons; one
            percent may contain premature stop codons; five percent may have
            frame shifts in a coding region. A sequence is considered to be
            5'-truncated if it lacks the translation initiation start (ATG). A
            C-terminal end of the encoded protein. Please note that these cDNA
            sequences are derived from the Ws or Laer ecotypes and therefore
            may contain polymorphisms when compared to sequences from Col-0.

```

```

FEATURES             source
Location/Qualifiers
     1..1541
     /organism="Arabidopsis thaliana"
     /db_xref="taxon:3702"
     /clone="100372"
     59..1459
     /codon_start=1
     /product="vacuolar processing enzyme/asparaginyl
     endopeptidase, putative"
     /protein_id="AA060827.1"
     /db_xref="GI:21536495"
     /translation="MSSPLGHFOILVFLHALIFSAESRRKTQLLNDNDVSSDKSAKG
     TRVALVAGSNEYNYRHQADICHAYQILRKGLKDENIIVFMYDDIARSENPRPGV
     IINKPGEDVYKGVKPKDYKAEVNVONFYNVLLGNESGVTGGNGKVVKSGPNDFIY
     YADHGAPGLIAMPTGDEVAKDFNEVLEKMKRKIKYMKVIYVEACESGSMFGILAK
     NLNIYAVTAANSKSSWGVSYPPEPGTICLTGDTFNSIWSLESDSLDMSKEFL
     EQQYHVVKRRVSDVPETSHVCRFEGTEKMLKDYLSYIGRNPNNDNFTFTFESSPLS
     NSGLVNPIDPLLYLQKKIQKAPMGSLEKQAKKLIDENKRNKQIDQSIDLRLSV
     KOTNLNLILTSTRTTGTGLOPLVDDMDCFKTLVNSFKNHCATVHYGLKYTGALANICNG
     VDYKQTVSAIEQACSM"

BASE COUNT  499 a 287 c 321 g 434 t
ORIGIN
Alignment Scores:
Pred. No.:      1,52e-207      Length:      1541
Score:          2446.00        Matches:      465
Percent Similarity: 99.79%      Conservative: 0
Best Local Similarity: 99.79%      Mismatches: 1
Query Match:      99.67%      Indels:      0
DB:              8            Gaps:          0

US-09-934-066-2 (1-466) x AY084227 (1-1541)

Qy      1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db      59 ATGTCTAGTCTCTTGGTCACCTTCAGATCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 118
Qy      21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAspLys 40
Db      119 TCAGCTGAGTCCGCAAAACCCCAATTGCTGAACGATAATGATGTTGAATCTAGGCAAG 178
Qy      41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db      179 AGTCAAAGACACACAGCATGGCTGTTTGTCTGGATCAATGATATATACTAC 238
Qy      61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAsp 80
Db      239 AGGCATCAGGCTGACATATGCCACGCGTATCAGATACCTCGAAAAGCGGTTTAAAAGAT 298
Qy      81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db      299 GAAACATCATCTGTTTGTATGATGATATCGCTTCTCGGAGAAATCTTAGGCCT 358
Qy      101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db      359 GGAGTTATCATTAATAAACCCAGATGGAGAGAGATGTTTATAAGGAGTTCCTAAGGACTAC 418
Qy      121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db      419 ACTAAAGAGCTGTTTATGTGTTTAACTTCTACAAATGTTTACTTGGAAATGAAAGTGC 478
Qy      141 ValThrGlyGlyAsnGlyLysValValLysSerClyProAsnAspAsnIlePheIleTyr 160
Db      479 GTCAAGAGGAAATGGCAAAAGTTGTGAAAGTGGTCTTAATGATAATATCTTCATCTAT 538
Qy      161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db      539 TATGCTGACCATGGAGCTCTCTGGCTTAATAGCATGCCACCTGCTGATGAAGTATGCA 598
Qy      181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200

```

Genset carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.

```

Db      599  AAAGATTTCAATGAAGTCTTGAGAAGATGCTAAGAGAAAATATACAAAGATGGTG 658
Qy      201  ILETYRValGluAlaCysGluSerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db      659  ATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTCAAGGGATTTTAAAGAAAATCTC 718
Qy      221  AsnIleTYRAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTYRcysPro 240
Db      719  ACATATACCGAGTACGCTGCTCTTCTGAGATGGAATCTCTCGCGGATACATTAACATC 778
Qy      241  GluSerTYRProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db      779  GAGTCATATCCCTCCCTCTCTGAGATGGAATCTCTCGCGGATACATTAACATC 838
Qy      261  SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTYR 280
Db      839  TCTTGCTGTGAGCAGAGTGCCTTCATGACATGACCAAGAGACTTTGGAGCAACAATAC 898
Qy      281  HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db      899  CAGCTTGTAAAGAGAAGTAGGATCTGATCTACACAGACTTCTCATGTATGCCGTTTC 958
Qy      301  GlyThrGluLysMetLeuLysAspTYRLeuSerSerTYRileGlyATGAsnProGluAsn 320
Db      959  GGAACAGACAGACATGCTTAAGATTAATCTTCTTACATTTGGAAGAAATCTGAAAC 1018
Qy      321  AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db      1019  GATACTTACCTTTCACCGAATCTTTCTCCACCAATCTCTAATCTGCTGGCTGGCAAT 1078
Qy      341  ProArgAspIleProLeuLeuTYRLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db      1079  CCGCGGGATATTCTCTGCTATACCTCAGAGAAAGATTCAAAAGCTCCAAATCGGATCA 1138
Qy      361  LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db      1139  CTTGAAGCAAGAAAGCTCAGAAGAAATTTGCTTACGCAAAAGATCATAGAAACAATC 1198
Qy      381  AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db      1199  GATCAGAGCATTTACAGACATTTCTGGCGCTTTCAGTTAAACAACAATGCTTAAATCTC 1258
Qy      401  LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db      1259  TTAACCTTCACAAAGAACACAGGACAGCTCTTGTAGACGATTGGGATTCCTCAAGACT 1318
Qy      421  LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTYRLeuLysTYRThr 440
Db      1319  CTAGTTAATAGCTTCAAGAAACACTGCGTGCACAGCTGATACGGAATGAAGTATACA 1378
Qy      441  GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db      1379  GGAGCGCTTGGCAATATCTGCAATATGGAGTGGATGGAAGCAAACTGTTTCAGCCATT 1438
Qy      461  GluGlnAlaCysSerMet 466
Db      1439  GAACAGCTTGCTCGATG 1456

RESULT 3
AP000383
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (sites)

```

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

exon

CDS

CDS

Kaneko, T., Katoh, T., Sato, S., Nakamura, A., Asamizu, E. and Tabata, S. Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC DNA Res. 7 (3), 217-221 (2000)

20363099

2 (bases 1 to 83253)

Kaneko, T., Katoh, T., Sato, S., Nakamura, Y., Asamizu, E. and Tabata, S. Direct Submission

Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp). Tel: 81-438-52-3935, Fax: 81-438-52-3934

Address for correspondence: kaos@kazusa.or.jp

For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MAL21

Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.

The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/Grail-1.3/>), GENSCAN (Chris Burge, MIT, <http://CC8-081.mit.edu/GENSCAN.html>), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://gremmlin.zozi.lastate.edu/cgi-bin/sp.cgi>).

Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).

This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MZE19 and the 3' clone is MOC12.

Location/Qualifiers

1. 83253

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/chromosome="3"

/clone="MAL21"

/clone_lib="Mitsui P1"

/complement(1..376)

/product="nucleoid chloroplast DNA-binding protein-like"

/notes="CDS is reported in Acc# AP002050"

gene_id="MZE19.7"

/number=1

/evidence=not_experimental

/complement(join(4273..4315,4392..4510,4727..4902,5273..5377,5836..5914,6019..6096,6222..6290,6555..6620,7231..7266,7418..7486,7659..7724,7870..8163))

/note="gene_id:MAL21.2"

/codon_start=1

/evidence=not_experimental

/product="protein arginine N-methyltransferase-like protein"

/protein_id="BAB01859.1"

/db_xref="GI:9293956"

/translation="MOSGGDFSNFGPHDHHRELEDKQPSLSFSGRAKRSIAGAR DPRGLANLVRVSDGLGEHKSLSESPPTCTDFDVAVFHYAHVTHEEMIKDRAR TETRYREALIMOHOSLIEGKVVVDVGGCTGTLSIFCAOAGAKRVYAVDASDIADQAEVY KANGLSKDVLVLRHVEDVEIDVDVYIIESEVMGYMLLYESMLGSLVITARDMLKPG LILPLSHATLYMAPISHPDRYSHSIDFWRNVYGDMSAMMOLAKOCAPPEPSVSGE NVLTWPVEVPEFSGPASPAPKNTSETSIASGSSISPSGVEVQKKTNPDSALVLS TSPSPPTTHOOTIVFYDPIEDVDQDVTEGVSVTLSQSKENKRMNTHLEYSSAGRSF VKESVVR"

10472..11680

/note="emb|CAB878723.1"

gene_id:MAL21.3

/similar to unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB01860.1"

/db_xref="GI:9293957"

/translation="MTMMSDLSQLLEEILSRVPTSLGAVRSTCKRWNTLFKDRILC
 KACETDQFRFIMKKYKLSMRFDLNGTNEDEGTEFVDPISIRELGHFFNQVTKSVF
 QCDGLLVCRLVWVWVNLGOIRWETESGNTNRLFDVAIGYDNNRKHKILRF
 LEYNNMKHRELEYEYDFSSNSWRLDIAPWEIESYORGASLKGNTYFIAKEKTE
 YEEDGEPEDANLLCFDFTESGFQFLPLFPQHYLDVGALSGLGDEKLAALFOCGD
 TDUSEVILWTTLTETNTVSNWNLPLKVDMEPHYGRSMFDYIGSGFIDEKELAVVF
 QFDSGIRVEDATYIIGENGYKVKRLGAPANOQGCFFSPVCSFYVSLVQINOI
 VGFKKEREET
 join(15535..15806,16005..16155,16253..16435,16674..16748,
 16945..17100,17229..17327,17425..17552,17741..17819,
 17917..18282)
 /note="gene_id:MAL21.4"
 /codon_start=1
 /evidence=not_experimental
 /product="chloroplast outer envelope hexokinase"
 /protein_id="BAB01861.1"
 /db_xref="GI:9293958"
 /translation="MGKVLVLTAAAVACSVATVVRMRKGRKRVRVGLKKDL
 EACETPLGRKQWDAIAVEMQAGLVSEGSKLMLTETVDDLPGNSGTGTYYALHL
 GGYSFIRIKVHLGQSRSLSEVQDVERHSIPTSLMNSTSEVLFDFLASSLQRTIEKEN
 DFLSQPLKRELAFTFSPVKQTSISSGVILKWTGPAISEMAGEDAECLQALNKR
 GLDIRAALVNDTVGLSGFGHFDPTIAAVFGTSGNACVLERTDAIKCONPRTTS
 GSMVNMENGFNSRLPRTSYDLEDAESMNSNDMGFEKMGIMGLGDIVRVILKM
 SQESDIFGPSSILSTFVLRTNSVSNHEDDISELQEVARIKLDGSEVPWKVRKL
 VVKCDVVTAAALAAAGIAGILKVKYGRDGGSGGRSDKOIMRRTVVAVEGGLYLN
 RMFREYDEALRDLGEDVAQHVVVKAMEDSGSSIGSALLASSOSVQOTIPSV"
 complement(join(18762..18845,18933..19038,19118..19221,
 19306..19434,19520..19624,19780..19914,19994..20080,
 20208..20304,20550..20767,20863..20928,21116..21211,
 21364..21414,21501..21560,21669..21755,21869..21957,
 22170..22234,22412..22484))
 /note="gene_id:MAL21.5"
 /codon_start=1
 /evidence=not_experimental
 /product="chaperonin, t-complex protein alpha subunit"
 /protein_id="BAB01862.1"
 /db_xref="GI:9293959"
 /translation="MSTSAQNPDISGRQSDVYRTQNVMACQAVSNIVKTSIGPVGL
 DKMLDDIGDVTITNDGATILRMEVEHPAAKVLVELAEQDREVGGDTTSVIVAAE
 LLKRLANDVRNKHPTPTISGYRLAMRECKYIEEKLVTKVEKLGKVPILNCAKTSWS
 SKLISGSDSDFANLVAVSVKMTNORGEIKYPIKINTILKAHGOSARDQSYLLNGA
 LNTGRAAGWPLRVSPAKIACLOPNQKTKMQLGVQVAVNDPRELEKIRQREADMTKE
 RIEKLKAGANVILTKGIDDMALKYFEAGALAVRRKEDMRHAKATGATITVTF
 ADMGEETFOPHLGSADVEVERIADDDVILKGTSTSAVSLIRGANDYMLDENE
 RALHDALCIKVRLESTNVAGGAVESALSVEHLIATLIGSREQLIAIEFADALI
 PKVLAYNAKDKATELVAKLRVYHHTAQTAKDKKHYSKRKKSGLDLVNGTIRNNLE
 AGVTEPAMSKVKIIOFATEAATILRIDDMIKLVKDESGQEE"
 complement(join(23145..23191,23289..23348,23451..23514,
 23623..23850,23934..23961,24383..24501))
 /note="gene_id:MAL21.6"
 /codon_start=1
 /evidence=not_experimental
 /product="ubiquitin conjugating protein-like"
 /protein_id="BAB01863.1"
 /db_xref="GI:9293960"
 /translation="MATVNGYGTNPAAATTPAATGSKQSPAPPTKVDHSHVLRKLOSE
 LMGLMGADGFIISAFPEEDNIFCWKGTITGSDFTEGTEYRLSTFSDNYPKSPKV
 KFETCCFHPNVVDYGNICLDIOLQDKWSSAYDVRTILLISQSLGEPNISSPLNQAQ
 LWSNOEYRKWKVEKLYPLNA"
 complement(join(25168..25258,25476..25644,25764..25981,
 26173..26234,26335..26425,26620..26689,26796..26843))
 /note="gene_id:MAL21.7"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01864.1"
 /db_xref="GI:9293961"
 /translation="MEALYAKLYDKYTKLQKKYSEYDEINKQEOKFLFTVSAESEL
 MEHLRGENSLSEVKEKRLNEIISIRGRDDKLECKOKLLMEELVKNKLSSEVVKIK
 ELVQEHPRNVDOQSKKORRKPESARVTRSMIKRSLSEDLVETDVMYSPDISHKH
 KAKEPLLVPOCCRTYDGSSSASCTFOALGKHLGLMKLSTNNKGRACIVASHT
 TGLSFLSTFINNPGSESELLYPASLGTFQRVAPDEWREVIKFTSMKRCIFFEVSFR
 VIKLNC"

27734..27973
 /note="contains similarity to cytochrome P450
 gene_id:MAL21.8"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01865.1"
 /db_xref="GI:9293962"
 /translation="MMLQCEDWTKTKGDKVMEAIRGLNLTMAHPLKCIPIVARNLNL
 ACSNLKPCNEFAKLLVRGSSAPLCMFPCPSLPV"
 join(29031..29951,30031..30681)
 /note="gene_id:MAL21.9"
 /codon_start=1
 /evidence=not_experimental
 /product="cytochrome P450-like protein"
 /protein_id="BAB01866.1"
 /db_xref="GI:9293963"
 /translation="MAAMNIVDFQNCFFVLLSLFLSLCYLSLFRKPKERLHYDLPP
 SPPTPIIGHLLHLLSVLRHSIQTLTKYGSILYLRVFSFPVALVSASAIYEIFRE
 HDNISFRPPPTDDSLFAGSFSTSAQYDKWMKKLLVTLNLGPOALERSGFRA
 DELDLFTENLLDKMKKESVDICVEALKLSNNSTCKIMGRSCSENGEAEVRALAT
 QGLDTKILLANMLRAGFKLVVSLFRKEMDVSRFDELLERILVHEHDKLDMHQ
 GTDLVALLAACRDKNAEYKISRHHIKSFEADLLFASTDTFVOTQWVAELINPNV
 LERLCGLDSVVGKARLIQETDLPNLPLYLQAVYKEGLRHPGLPFAFSEQEGRIGG
 FYPEKTLTMINAYVMRSDSDWEDPDEFKPERFLASSRSEQEKERQALKIYAFGS
 GRSCPQENLAYIFLGTAGVMVQGFEMRIKEKVNMEANVGLSLTMYPLKVTVP
 RTLVPITQNPINRSS"
 join(32158..32475,32844..33086,33190..33813)
 /note="contains similarity to cytochrome P450
 gene_id:MAL21.10"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB01867.1"
 /db_xref="GI:9293964"
 /translation="MAAMINVEFQNCFFIMLCIFSLCYLSLFRKPSRONSVDCL
 PPSPSPPIIGHLLHLLSVLRHSQFQKISSKYGLLHLRFPNVPILVVSASVAYDIP
 RVHDKFLAATLRLLEKLGISULFQKEMVSGRFGDELLERILREHEEKDEHDDTOM
 MDALAYNDKAEYKTRTQIKAFIVDMFIAGTDSALTSQTQMAEIIINNPNVRI
 REKIDSVVGSRILQETDLPKLPYQAVVKEGLRHPPTPLMVRFOEGKVGKGYIP
 ASITLVVNGVAVMRDPNVWEPEKPERFLASSRLEEEIIEQALKYIAFGSGRG
 CPGTNIAYIFVGTAGIMVMVCFDWMKIKGDKVMEKALGGLNLTMAHPLKCTPVARS
 PF"
 join(34864..34912,34957..35241,35310..35443,35870..36133)
 Alignment Scores:
 Pred. No.: 2,32e-182 Length: 83253
 Score: 2188.00 Matches: 463
 Percent Similarity: 67.89% Conservative: 0
 Best Local Similarity: 67.89% Mismatches: 3
 Query Match: 89.16% Indels: 219
 DB: 8 Gaps: 7
 US-09-934-066-2 (1-466) x AP000383 (1-83253)
 Qy 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
 Db 72700 ATGCTAGTCTCTCTGTGTCACCTTTTCAGATCTTTCTTCATGCTTTGCTTATCTC 72759
 Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspValGluSerSerAspLys 40
 Db 72760 TCAGCTGAGTCCCGCAAAACCCCAATTCCTGAACATATGATGTGATCTAGGCACAG 72819
 Qy 41 SerAlaLysGlyThrArgTyrAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
 Db 72820 AGTCAAAAGGCACACGCTGGCTGTTTATGTTGCTGATCAATGAATATATACTAC 72879
 Qy 61 ArgHis----- 62
 Db 72880 AGGCATCAGTTGTGTTAAATTTATGTTTGAAGCTTTAACAACAAAAAGGTCCTA 72939
 Qy 63 -----GlnAlaAspIleCysHis 68
 Db 72940 AGCGAGATTTGTATGAACATAATCGACCGACGCTTTTATTTTCACAGGCTGACATATGCCAC 72999

QY	69	AlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMetTyr	88
DB	73000	GCCTATCAGATACTCCGAAAGCGGTTTAAAGATGAAACATCATTTGTGTTATGTAT	73055
QY	89	AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp	108
DB	73060	GATGATATCGCGTTTCTCGAGAAATCCTAGGCCTGAGTTATCATTAATAAACAGAT	73119
QY	109	GlyGluAspValTyrLysGlyValProLys	118
DB	73120	GGAGAAAGATGTTTATAAAGGAGTCCCTAA - GGTTCCTTATTTCTACTCTCTTTTGTGCGTTA	73178
QY	118	-----	118
DB	73179	TTTCTACGTTGAATTCAAATACATATATATATTCAAAGTTTGTGTTGTTGTTGGTAG	73238
QY	119	AspTyrThrLysGluAlaValAsnValGluAsnPheTyrAsnValLeuGlyAsnGlu	138
DB	73239	GACTACACTAAAGAGCTGTTAATGTTCAAACTCTACAACTGTGTACTTGGAAATGAA	73298
QY	139	SerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePhe	158
DB	73299	AGTGGCGTCACAGAGGAAATGGCAAAAGTTGTGAAAGTGGTCTTAATCATATATCTTC	73358
QY	159	IleTyrTyrAlaAspHisGlyAlaProGlyLeuIle	170
DB	73359	ATCTATTATGCTGACCATGGAGCTCTGGCTTAAATAGGTTTTCTTAATTTATGAATTA	73418
QY	170	-----	170
DB	73419	TTACGTACCATCAATCCATATCTATATATAAAGATTTTCTCTGTGATACTACGAAACCGCG	73478
QY	171	-----AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeu	187
DB	73479	ATTTTCTCAGCGATGCCCATCTGGTGATGAAGTTATGCGAAAGATTTCAATCAAGCTCTG	73538
QY	188	GluLysMetHisLysArgLysLysTyrAsnLysMetVal	200
DB	73539	CAGAAGATGCAATAGAGAAAAAATACAAAGATGGT - ATATAACTCAACCATTCGTTA	73597
QY	200	-----	200
DB	73598	CCTAGCTTTATACATATGTGTTCTGTTTTGAATCTCTATGCTGTGTTTTTTGGATGTT	73657
QY	201	-----IleTyrValGluAlaLysGluSerGlySerMetPheGluGlyLeuLysLys	218
DB	73658	TAGCTGATCTATGTTGAAGCATGTGAATCAGGAAGATGTTTGAAGGATTTTAAAGAA	73717
QY	219	AsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyr	238
DB	73718	ATCTCAACATATACCAGTACTCTGCTAATTTCAAAGAGACAGCTGGGGAGTTTAC	73777
QY	239	CysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPhe	258
DB	73778	TGTCCTGAGTCATCTCCCTCCTCTCTCAGATTTGGAACCTTCTCTCGCGCATACATTT	73837
QY	259	SerIleSerTrpLeuGlu	264
DB	73838	AGCATCTCTGGCTTGAGGACAGCTACTGCAACAAAAAGATTCAATCTTATGGACTA	73897
QY	264	-----	264
DB	73898	TTGCAATGATTTGATTTGTTCTTGAGAAATATTTGTTCAATTTGTTCTATGTTTGTGTT	73957
QY	265	-----AspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisVa	282
DB	73958	GTTTGGGACAGTGACCTTCATGACATAGCAAGAGACTTTGGAGCAACAATACCACGT	74017
QY	282	IlyalLysArgValGlySerAspValProGluThrSerHisValCysArgPheGlyPh	302
DB	74018	TGTNAAGACAGAGTAGGATCTGATGATACAGAGACTTCTCATGTATGCGTTTCGGAAC	74077
QY	302	rGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAs	322

Db	74078	AGAGAAGATGCTTAAGATTAATCTTCTCTTACATTTGGAAGAAATCTCTGAACACGATA	74137
Qy	322	nPhetrPhetrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProA	342
Db	74138	CTTCACCTTTCACGAATCTTTCTCTCACAATCTCTAAATCTTGGCTTGCTCAATCCGCG	74197
Qy	342	gAspIleProLeuLeuTyrLeuGlnArgLysIle-	353
Db	74198	CGATATTCTCTGTATACCTCCAGAGAAAGGT-GAGCTTTTTCGGGTTTTTTCATCAT	74256
Qy	354	-----	365
Db	74257	TTTAAACGAAGAGATTTTCAGCATGTTTTAATGTTTATTTCATCTCTTAGATTTCAAAAAGC	74316
Qy	356	aProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnH1	376
Db	74317	TCCAAATGGGATCACTTCAAGAACGAAGAGCTCAGAGAAATTCCTTGACGAAAGAATCA	74376
Qy	376	sArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAs	396
Db	74377	TAGGAACAAATCGATCAGACATTCACAGACATCTCGCGCTTTCAGTTAAACAAACCA	74436
Qy	396	nValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTTPAs	416
Db	74437	TGCTTTAAATCTCTTAACCTTCCAAAGAACACAGACAGCGCTCTGTGTAGACGATTGGGA	74496
Qy	416	pCysPheLysThrLeu-	421
Db	74497	TTGCTTCAAGACTCTAGTAACAAACCCACATCTCAAAACCTTGTTACTTGTCTTACGCA	74556
Qy	422	-----	433
Db	74557	CAACCATTCATTTACTTAAACCAACAGTGTATATCAATGAAATCGCAGGTTAATAGCTT	74616
Qy	425	eLysAsnHisCysGlyAlaThrValHisIstIrglyLeuLysTyrThrGlyAlaLeuAlaAs	445
Db	74617	CAGAATCACTCGGGTGCACCGTGCATTCAGGATTTGAAGTATACAGAGAGCGCTTGCCAA	74676
Qy	445	nIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSe	465
Db	74677	TATCTGCAATATGGGAGTGGATGTGCAAGCAAACTGTTTCAGCCATTCAACAAGCTTTGTTT	74736
Qy	465	rMet 466	
Db	74737	GATG 74740	
RESULT 4			
LOCUS	D89971		
DEFINITION	Vigna mungo mRNA for asparaginyl endopeptidase (Vmpe-1), complete cds.	1724 bp	linear
ACCESSION	D89971		
VERSION	D89971.1	GI:4589395	
KEYWORDS	asparaginyl endopeptidase (Vmpe-1).		
SOURCE	Vigna mungo seedling cotyledon cDNA to mRNA, clone:pPEA.		
ORGANISM	Vigna mungo		
REFERENCE	1		
AUTHORS	Okamoto,T. and Minamikawa,T.		
TITLE	Molecular cloning and characterization of Vigna mungo processing enzyme 1 (Vmpe-1), an asparaginyl endopeptidase possibly involved in post-translational processing of a vacuolar cysteine endopeptidase (SH-EP)		
JOURNAL	Plant Mol. Biol.	39 (1), 63-73	(1999)
MEDLINE	99178794		
REFERENCE	2		
AUTHORS	Okamoto,T.		
TITLE	Direct Submission		
JOURNAL	Submitted, 1999		


```

Db 808 TATAGCACITGTTGGTGACTGTATAGTATTGCTTGGATGGAGCAGTGCACATAC 867
Qy AspNetSerLysGluThrLeuGluGlnThrHisValValLysArgArgValGlySer 289
Db 868 AACTTGGGACAGAAACTCTTCCACGACAGTATGAATGGTTAAACAAAGACTGCTAGT 927
Qy AspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyr 309
Db 928 TACAATTCGTATGCTCTCATGTCATGCAATATGGTATATCGGCTTAGCAAGAACAAT 987
Qy LeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPhe 329
Db 988 CTCTTACGCTACTGGGTACAAATCTGCAATGATACATACCTTTTGGTGGAGAAC 1047
Qy Ser---SerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLysTyr 348
Db 1048 TCCTTAAGCCACATCTAAAGCT-----GTTAACCCAGGAGATGCCGATCTCTTGAT 1101
Qy LeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLys 368
Db 1102 TTCTGGGACANGTACCGAAGGACCTGAGGCGCACTCTTAGAAGGCTGAAGCTCAGAAG 1161
Qy LysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeu 388
Db 1162 CACTTTTGAAGCAATGCTCATAGAAATGCTAGATGACACACAGCATCAAACTTATTGGA 1221
Qy ArgLeuSerValLysGlnThrAsnValLeuAsnLeuLysSerThrArgThrThrGly 408
Db 1222 AACCTCTTATTTGGAATTGGAAGGTCAGAGATCTTGAACACATGTTCCGACCTGCTGT 1281
Qy GlnProLeuValAspAspTyrAspCysPheLysThrLeuValAsnSerPheLysAsnHis 428
Db 1282 CAGCCTCTTGTGATGACCTGGGCTGCCCTCAAAATCACTGTTGAGTGGCAGC 1341
Qy CysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsn 448
Db 1342 TGTGGAGCTCTATCGCAATATGGAATGAGACACATGCTCTCTTGGCAACATCTGTAAT 1401
Qy MetGlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCys 464
Db 1402 ACAGGAATTTGGGAAGAAGATGGCTGGAAGCATCAGCACAGCTGT 1449

RESULT 7
AF238384 1576 bp mRNA linear PLN 02-MAR-2001
LOCUS Vigna radiata asparaginyl endopeptidase (PEI) mRNA, complete cds.
DEFINITION AF238384
ACCESSION AF238384
VERSION AF238384.1 GI:13183094
KEYWORDS Vigna radiata.
SOURCE Vigna radiata
ORGANISM Vigna radiata
REFERENCE 1 (bases 1 to 1576)
AUTHORS Yang,W.H. and Chen,C.S.
TITLE Vigna radiata mRNA for asparaginyl endopeptidase, complete cds
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1576)
AUTHORS Yang,W.H. and Chen,C.S.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-2000) Institute of Botany, Academia, Sinica, Nankang, Taipei, Taiwan 11529, Republic of China

FEATURES
source
1..1576
/organism="Vigna radiata"
/cultivar="Tainan 5"
/db_xref="taxon:157791"
/tissue_type="cotyledon"
/dev_stage="germinating seeds"
1..1576
/gene="PEI"

gene

```

```

CDS
19..1470
/contig="pe1"
/codon_start=1
/product="asparaginyl endopeptidase"
/protein_id="AAK15049.1"
/db_xref="GI:13183095"
/translaton="MATTPTSLTFLPLATVALVAAGRDLYGDFLPLPSDNGDDN
VGTWATILFAGSNGVWYRHQADICHAYOILKGLKEERILVPMYDDIAPNDNPR
PGVINKPGDDVYGVPKDYCEDATAHNFYSALLGSKSALTGSGKAVNSGDDRI
FTEYDHGCGPVLYASDLVULKKHASCYKLVLYLEACAGSLFEGIL
LEPDINIYATASNAEESNGHTYCPGEPSPPEYSTCLGDLISVAMEGDSRINLRT
ESHQYTKVKDRTLSGGHTGSHVQGTGVFERSKDTLFLYLGTDLPANDNLTFVDENSL
WSSSTNQRDADLVHFHWHKRAPEGSPKNEARKQVLEVMSHRMHLDDSVKLVGKL
LGFKEAPVLANRPAAGSLVDDWACLVKRVTFETHCGSLSQYGMKMRMSFANIC
VGKIKKQMEASAQAQCVTPASSWSLQGFSA"
BASE COUNT 419 a 327 c 377 g 453 t
ORIGIN
Alignment Scores: 6.97e-102 Length: 1576
Pred. No.: 1254.50 Matches: 231
Score: 70.73% Conservative: 71
Percent Similarity: 54.10% Mismatches: 124
Best Local Similarity: 51.12% Indels: 1
Query Match: 8 Gaps: 1
DB:
US-09-934-066-2 (1-466) x AF238384 (1-1576)
Qy 38 SerAspLysSerAlaLysGlyThrArgTyrPalaValLeuValAlaGlySerAsnGluTyr 57
Db 139 AACGATGACAACTCAAGGGGACGCGCTGGCGCATCTCTTCGCGGTTCCGAACGGCTAT 198
Qy 58 TyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGly 77
Db 199 TGGAACTACAGACATCAGCGCCGATATTTCATGCCTCAANTACTCAGGAAGCTGCC 258
Qy 78 LeuLysAspGluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsn 97
Db 259 CTGAAGAAGAAATATCATTTGTTATGATGATGATGATGATGATGATGATGATGAT 318
Qy 98 ProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValPro 117
Db 319 CCAAGCCCTGGAGTCATCATTAACAAACGACGGGATGATGATGATGATGATGATGAT 378
Qy 118 LysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsn 137
Db 379 AAGGATTACACTGGCGAAGATGCTACTGCTCATAACTTTTATCTGCTTTACTTGGAGAT 438
Qy 138 GluSerGlyValThrGlyLysValLysValLysValLysValLysValLysValLys 157
Db 439 AAGTCGCGCTTACCCTGGCGTGGCAAGGTTGTGAACAGTGGCGCGCATGACCGCATA 498
Qy 158 PheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGlu 177
Db 499 TTTATATTCTACTCTGATCATGCTGCTCCAGGGGTTCTCGGACGCGCTGCTGCTTAC 558
Qy 178 ValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsn 197
Db 559 ATATATGCTCTCATTTGCTTGAAGTATTGAAGAAAGAAAGACGCTTCTTGAACGATATAA 618
Qy 198 LysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyLysLys 217
Db 619 AACCTAGTATTTTATTTGAGCGCATGCTGAAGCTGGAAGTATTTTGAAGGCTCTCTCT 678
Qy 218 LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTroglyVal 237
Db 679 GAAGATATCATATTTATTCACACACCGGCTTCGATGCGGAGAAAGTAGTTGGGACACA 738
Qy 238 TyrCysProGluSerTyrProProProProSerGluIleGlyThrCysLeuGlyAspThr 257
Db 739 TATTGCTGGGGAATATCTAGTCTCCCGCAGAAATATTCAACCTGCTGGGTGGGACCTG 798
Qy 258 PheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGlu 277

```


Db	550	CCAGGGGTTCTCGGACGCGCTCGTCTTACATATATGTCATCTGATTGGTTGAAGTA	609
Qy	187	LeuGluLysMetHisArgLysLysLysValIleTyrValGluAlaCys	206
Db	610	CTGAAGAAAACCATGCTTGGACGATATAAAACCTAGTATTTTATTTGGAGCATGT	669
Qy	207	GlusSerGlySerMetPheGluGlyLysLeuLysLysAsnLeuAsnIleTyrAlaValThr	226
Db	670	GAAGCTGGAGATCTTCTGAAGTCTCTCTCTGAAGATATCAATATTTATGCAACACG	729
Qy	227	AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProPro	246
Db	730	GCTTCGAATCCCGAAGAAGTAGTTGGGAACATATCTCTGGGAGTATCTTCAGTCCT	789
Qy	247	ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer	266
Db	790	CCCCAGAATATTCACCTGCTGGGTACACCTGATGATGCTGGTGGATGGAAGACGT	849
Qy	267	AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArgArg	286
Db	850	GATAGACACAACTTCGACAGCAATCTTTACACCAACAATATAAGTGGTGAAGATAGG	909
Qy	287	ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu	306
Db	910	ACTCTATCTGGAGGTGGTGTCTCAGCTGATGATGCTGATGATGATGATGATGATG	969
Qy	307	LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr	326
Db	970	AAGGATGCTCTCTCTCTGATTTGGGTACAGATCTGCAAAATGATAATTTGATCTTCGT	1029
Qy	327	GlusPheSerSerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeu	346
Db	1030	GATGAA--AACTCTTATGTCATCTCACTGATGATGATGATGATGATGATGATGATG	1086
Qy	347	LeuTyrLeuGluArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAla	366
Db	1087	GTCCATTTTGGCATAGTTCGCAAGCTCGGAGGTTCTCTAGAAAATGAAGCT	1146
Qy	367	GlnLysLeuLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAsp	386
Db	1147	CGAACAAGTCTCGAAGTATGCTCAGAGATGATATAGACGACAGTGAAGAACTT	1206
Qy	387	IleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuSerThrArgThr	406
Db	1207	GTGGAAAGCTTTTATTTGGATTTGAAAAGGCTCTGAAAGTACTGAATGCTGTGAGACCG	1266
Qy	407	ThrGlyGlnProLeuValAspTrpAspCysPheLysThrLeuValAsnSerPheLys	426
Db	1267	GCTGGATCGGCACCTTTGTCATGACTGGCCCTCGCTGAAAACCATGCTTAAGCACTTTGAG	1326
Qy	427	AsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIle	446
Db	1327	ACACATTTGTCATCTGCTCAGTATGGATGAACACATGATGCTCTTTCGCAATATC	1386
Qy	447	CysAsnMetGlyAlaValLysGlnThrValSerAlaIleGluGlnAlaCys	464
Db	1387	TGCAATGTAGGAATAAAGAAAGACGAATGCTGAAGCCTCAGCAGCAAGCTTGT	1440
RESULT	9		
LOCUS	AY133531	1485 bp	mRNA
DEFINITION	Arabidopsis thaliana At4g32940/F26P21_60	linear	PLN 07-AUG-2002
ACCESSION	AY133531		complete cds.
VERSION	AY133531.1	GI:22137031	
KEYWORDS	FLI_CDNA...		
SOURCE	Chate cress.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.		
REFERENCE	1 (bases 1 to 1485)		
AUTHORS	Kim, C.J., Chen, H., Cheuk, R., Shinn, P., Banh, J., Bowser, L.,		
	Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,		

Ishida, J., Kamiya, A., Karlin-Neumann, G., Kawai, J.,			
Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,			
Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,			
Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K.,			
Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A.			
and Ecker, J.R.			
Arabidopsis ORF clones			
Unpublished			
2 (bases 1 to 1485)			
Kim, C.J., Chen, H., Cheuk, R., Shinn, P., Banh, J., Bowser, L.,			
Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,			
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,			
Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,			
Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,			
Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K.,			
Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A.			
and Ecker, J.R.			
Direct Submission			
Submitted (17-JUL-2002) Salk Institute Genomic Analysis Laboratory			
(SIGNAL), Plant Biology Laboratory, The Salk Institute for			
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,			
USA			
RIKEN Genomic Sciences Center (GSC) members carried out the			
collection and clustering of RAFL CDNAs (RAFL cDNA : 'RIKEN			
Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J.,			
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,			
Hayashizaki, Y. and Shinozaki, K.			
The Salk, Stanford, PCEC (SSP) Consortium members constructed and			
sequenced the PUNI (ORF) clones using the RAFL CDNAs : Kim, C.J.,			
Chen, H., Cheuk, R., Shinn, P., Banh, J., Bowser, L., Chang, E.,			
Dale, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B.,			
Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J.,			
Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,			
Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A.,			
and Ecker, J.R.			
Kim, C.J. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to			
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)			
contributed equally to this work as PIs.			
Location/Qualifiers			
1. 1485			
/organism="Arabidopsis thaliana"			
/db_xref="taxon:3702"			
/chromosome="4"			
/clone="U14418"			
/note="This clone is in pUNI 51"			
ecotype="Columbia"			
1. 1485			
/note="gamma-VPE (vacuolar processing enzyme)"			
/codon_start=1			
/product="At4g32940/F26P21_60"			
/protein_id="AA091361.1"			
/db_xref="GI:22137031"			
/translation="MATTMTRVSGVLFVLLSLVAVSAARSGPDDVILKPSASRF			
FRPAENDDSSNGTRWAVLVAGSSGYNRYHQADICHAYOLLRKGLKEENIVVFMVD			
DIANNYENPRCTIISPHGKDYGVKDYTDGDDVNDVNFVILGDKTKAVKGSCK			
VYDGPNDHIFLEYSNHCPCVLCMPSTSPYLXANDNDVLLKHALCTYKGLVYLA			
CEGSIFFCLLEGLNIYATTASNAEESMGTCGPEPSPPEYETCLDGLYSVAM			
EDSCHNLOTETLHQOYELVRRTPAPVCSYGVSHVMOYGVGISKNDLUDLYNGTNPN			
DNFTFADANSKPPSRVNTADLYHFWKTRAPGSARKTCAQKQVLEAMSHRLH			
INDSVILGKILFGLSGRGEVLNKRVSQGLPDVDMNCKNQVAFRHCGLSLSQYGI			
KIMRSFANICNAGIOMEAEASQACTTLPSTGPWSSLNRGFS			
BASE COUNT	409 a	308 c	365 g
ORIGIN			403 t
Alignment Scores:			
Pred. No.:	3.65e-101	Length:	1485
Score:	1246.00	Matches:	233
Percent Similarity:	69.89%	Conservative:	71
Best Local Similarity:	53.56%	Mismatches:	121
Query Match:	50.77%	Indels:	10
DB:	8	Gaps:	4

[illegible][illegible]

```

/Chromosome="4"
/clone="U10153"
/Note="This clone is in pUNI 51."
ecotype: Columbia
1..1516
/gene="At4g32940"
1..1485
/gene="At4g32940"
/codon_start=1
/evidence="experimental"
/product="putative vacuolar processing enzyme gamma-VPE"
/db_xref="GI:16323432"
/translation="MATTWIRVSGVVLVLLVSLVAVSARSPPDDVIKLPQASRF
FRPAEDDSDSGTWRWLVAGSSGYWNYRHOADICHAYQLLRGGGLKEENIVVFVYD
DIANYENPRGFTIINSHGKDVYGVPKDYTDNDVNLFAVILGDKRTAVKGGSGK
VYDSGNPHIFLFDSDGGPGVLPMTSPILYANDLVNKKHALGTYKSLVFLYEA
CESGIFLGLPELQNIYATTASNAEESWGTGCPPEEYETCLDGLYSVAMW
EDSGHNLQETLHQYLVKRTAPVGSYSHVMQYGDVSKDNLDMGTNPAN
DNFTFADANSLKPPSRVNRDADLVHFWERKAPESARKTEAQFVLEAMSHRLH
IDNSVLVKGILFGLISRGPEVLNKRVSAGPLVDVWNLCKNQVAFERHOGSLSQVGI
KMRSFANICNAGIQMEQMEPAASQACTTTLPTGPWSSLNRGESA"
1486..1516
3'UTR
/gene="At4g32940"
BASE COUNT 418 a 316 c 368 g 414 t
ORIGIN
Alignment Scores:
Pred. No.: 3,75e-101 Length: 1516
Score: 1246.00 Matches: 233
Percent Similarity: 69.89% Conservative: 71
Best Local Similarity: 53.56% Mismatches: 121
Query Match: 50.77% Indels: 10
DB: 8 Gaps: 4
US-09-934-066-2 (1-466) x AY059104 (1-1516)
QY 36 GluSerSerAspLysSerAlaValThrArgTrpAlaValLeuValAlaGlySerAsn 55
Db 145 GAAACGACGAGTATCACTCCCGTACTAGTGGGCTGTCTAGTCCGCGGATGAGC 204
QY 56 GluTyrTyrAsnTyrArgHisGlnAlaAspLysCysHisAlaTyrGlnLeuLeuArgLys 75
Db 205 GGATATTGGAATTCAGGCGATCAGGCTATATGCCATGCTATCACTTCTCGAGGAA 264
QY 76 GlyGlyLeuLysAspGluAsnIleValPheMetTyrAspPheAlaPheSerSer 95
Db 265 GTGGGATTGAAAGAGGAGCAATATGTTGTTATTCATGATGATGATGATGATGATGAT 324
QY 96 GluAsnProArgProGlyValIleIleAlaAsnLysProAspGlyGluAspValTyrLysGly 115
Db 325 GAGATCCAGGCGCTGGAACCATTTATCAACAGCCCTCATGGAAGAGATGCTATCAAGGA 384
QY 116 ValProLysAspTyrThrLysGluAlaValAlaAsnValGlnAsnPheTyrAsnValLeuLeu 135
Db 385 GTTCCCAAGGATATATATGAGAGATGATGATCAATGTTGATAATCTATTTGCTGTCATCCT 444
QY 136 GlyAsnGluSerGlyValTyrGlyAsnGlyLysValValLysSerGlyProAsnAsp 155
Db 445 GGAGACAAACTGCTTTAAAGGGGAGTGGGAGGTTGTGGATAGTGTGCTCTTATGAT 504
QY 156 AsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
Db 505 CATATCTCATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 564
QY 176 AspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLys 195
Db 565 CCTTACCTATATGCAATGATCTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 624
QY 196 TyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIle 215
Db 625 TATAAAGACTTGGTCTTTTATCTCGAAGCTTGCAGAACTGGAAGTATCTATTGGAAGGCTT 684

```

```

216 LeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrp 235
Db 685 CTTCTCTGAGGGTTTGAACATCTATGCCACAACATGTCATCAACGCGGAAGACAGTGG 744
QY 236 GlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGly 255
Db 745 GGTACCTATTGCTCGGAGAGAACCCAGCTCTCCACCGGAGTATGAACATCTTTAGGT 804
QY 256 AspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThr 275
Db 805 GACTTGTACAGTGTCTCTGGATGGAAGATAGTGTATGCACAATTTACAGACTGAGACT 864
QY 276 LeuGluGlnGlnTyrHisValValLysArgArg-----ValGlySerAspValPro 292
Db 865 CTGCACCAACAATATGAACCTTGTGAAAGAGGAGACTGCACCTGTGGTACTCTTAT--- 921
QY 293 GluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLysSerSer 312
Db 922 ---GGTCTTCATGCTATGCAATATATGCGATATAGGATATAGCAAGGATATATCTCGACTT 978
QY 313 TyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGlu-----SerPheSer 330
Db 979 TATATGGGAACAACCCCTGCCAATGCAATTTTACCTTTTGGGATGCGCAATCTCAATAAG 1038
QY 331 SerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGln 350
Db 1039 CCACCT-----TCAAGAGCTTACAAACAGCCCTGATGCGAGATCTTGTTCATTTTGG 1089
QY 351 ArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu 370
Db 1090 GAAAGTACCGAAGAACCCAGAGGTTCCAGCAAGAAAACAGAGCTCAGAACGACAGTA 1149
QY 371 LeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeu 390
Db 1150 CTGTAAGCCATGCTCACAGACTTCAATATGCAATACGCTGATGATGATGATGATGATGATG 1209
QY 391 SerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnPro 410
Db 1210 TTGTTTGGCATTTTCAGAGAGTCTCTGAAGTGTCTAAACAAAGTACGCTGCTGGGCACTT 1269
QY 411 LeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGly 430
Db 1270 CTAGTGTGATGCTGGAACCTGCTTAAATAATCAGGTGAGAGCTTTTCGAGAGGACCTGTGA 1329
QY 431 AlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGly 450
Db 1330 TCGCTGCTCTCAGTACGATATCAAGCACATGAGTCTTTTTCGAAACATCTGCAATCCAGG 1389
QY 451 ValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSer 465
Db 1390 ATTCAAATGGAGCAAAATGGAGGAGGAGCTTCACAGCGCTTGTACC 1434
RESULT 11
AF424619 1760 bp mRNA linear PLN 08-OCT-2001
Arabidopsis thaliana AT4g32940/F326P21_60 mRNA, complete cds.
AF424619
AF424619.1 GI:15983488
VERSION
FLI_CDNA
Arabidopsis thaliana
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1760)
Shinn, P., Chen, H., Cheuk, R., Kim, C. J., Koesema, E., Meyers, M. C.,
Ban, J., Bowser, L., Carninci, P., Dale, J. M., Goldsmith, A. D.,
Hayashizaki, Y., Ishida, J., Jiang, P. X., Jones, T., Kamiya, A.,
Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J. M., Lin, J., Liu, S. X.,
Miranda, M., Narusaka, M., Nguyen, M., Onodera, C. S., Palm, C. J.,
Pham, P. K., Quach, H. L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,
Tang, C. C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,
Shinozaki, K., Davis, R. W., Theologis, A. and Ecker, J. R.

```


US-09-934-066-2 (1-456) x PVZ99956 (1-1850)

Qy 10 IleLeuValPheLeuHisAlaLeuLeuLeuPheSerAlaGluSerArgLysThrGlnLeu 29
 Db 66 CTCCTCTCTCTCTGCGCTCGGCT-----TCTGCTGGCGGTGACCTC 110
 Qy 30 LeuAsnAsp-----AsnAspValGluSerSerAspLysSerAlaLysGly 44
 Db 111 GTCGGAGACTCTCCGACCTGCGCTGCTGATCCGGTAAGGGGAC---AACGTCACATGA 167
 Qy 45 ThrArgTrpAlaValLeuAlaValGlySerAsnGluTyrTyrAsnTyrArgHisGlnAla 64
 Db 168 ACGCGCTGGGCCATCTCTTCGCTGCTGCTGCGAGGGCTATTGGAAATPACAGACATCAGGCC 227
 Qy 65 AspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIle 84
 Db 228 GATATTTGTCATGCTATCAATTAATCTAGGAAAGGTGGCTGAAAGCATGAAATATCATTT 287
 Qy 85 ValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIle 104
 Db 288 GTTTTATGATGATGACATTTGATTCATTCATTCGAAATCCAAAGGGTGGAGTCATCATTT 347
 Qy 105 AsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAla 124
 Db 348 AACAGTCCAATGGAGATGAAGTTTATAAGGAGTTCCAAAGCATTTACACTGGTGAAGAT 407
 Qy 125 ValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGlyGly 144
 Db 408 GTTACTGCTCATAACTTTTATGCTGCTTACTTGGAGATAGTCGAAACTTACTGTGGC 467
 Qy 145 AsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHis 164
 Db 468 AGTGGGAAGTTGGATAGTGGCCCATGATGACATATTTATATCTACTCTCATCAT 527
 Qy 165 GlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsn 184
 Db 528 GTGGTCCAGGGGTGCTGGGCTCGCTGCTGCTGCTTACATTTATGATCTGCTCATGAAT 587
 Qy 185 GluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGlu 204
 Db 588 GAAGTATTTGAAGAAACAACTGCTCTGGAACCTATATAAACCTAGTATTTATTGGAG 647
 Qy 205 AlaCysGluSerGlySerMetPheGlyIleLeuLysLysAsnLeuAsnIleTyrAla 224
 Db 648 GCATGTGAATCTGAAGTATCTTTGAAGGTCTTCTTCCCTGAAGATATCAATGTTATGCA 707
 Qy 225 ValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrPro 244
 Db 708 ACAACAGCTTCGATGCGGATGAAGTAGTTGGGAACATATTTGCTGGGAGGATGCT 767
 Qy 245 ProProSerGluIleGlyThrCysLeuGlyAspThrPheSerTyrLeuGlu 264
 Db 768 AGTCTCCCCAGAAATATTCAACCTGCTTGGGTGACCTGTACAGTGTCTGGATGGA 827
 Qy 265 AspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnClnTyrHisValValLys 284
 Db 828 GACAGTATAGACACAAATTTGCCACAGAACTTTCACCAACAAATATAATTTGTTAA 887
 Qy 285 ArgArg---ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGlu 303
 Db 888 GAGAGGACTATATCTGAGGTTTATCTATGCTCTACGTGATGCGATGATGATGTA 947
 Qy 304 LysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPhe 323
 Db 948 GGGCTTACAGAGGATATCTCTCCATTTATTTGGGTACAGATCTTCAATGAGATTTG 1007
 Qy 324 ThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProArgAsp 343
 Db 1008 ACTTCTGTGATGAA---AACTCTTATGCTGATCTTCAAAAGCAGCTCAACCAAGGGAT 1064
 Qy 344 IleProLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSer 363
 Db 1065 GCTGATCTGCTCATTTTGGGATAAGTTCGCGAAAGCTCCTGAGGGTTCTCTTAAGAA 1124

Qy 364 LysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSer 383
 Db 1125 AATGAAGCTCGGAACAAGTTCTGGAAGTAATCTCTCACAGAAATGATATAGACGACAT 1184
 Qy 384 IleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSer 403
 Db 1185 GTGAAGCTTGTGGAAGCTTTTATTGTCATTTCAAAAGGCTCCTGAATTTACTGAATGCT 1244
 Qy 404 ThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThrLeuValAsn 423
 Db 1245 GTAAGACGGGTGGATCCGCACTGTTGATGACTGGGACTCCCTGAAACCACTGTTAAGG 1304
 Qy 424 SerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeu 443
 Db 1305 ACTTTTGAGACACATTTGTCATCTGTCACATGATGGATGAACACATGAGTCTCTTT 1364
 Qy 444 AlaAsnIleCysAsnMetGlyValAlaAspValLysGlnThrValSerAlaIleGluAla 463
 Db 1365 GCGAACATGTGCAATGTAGGAATAAGAGGACGAATGCTGCAAGCTCAGACCAAGCT 1424
 Qy 464 Cys 464
 Db 1425 TGT 1427

RESULT 14
 BVU309173 1980 bp mRNA linear PLN 23-MAY-2002
 LOCUS Beta vulgaris mRNA for putative vacuolar processing enzyme.
 DEFINITION AJ309173
 VERSION AJ309173.1 GI:14594818
 KEYWORDS vacuolar processing enzyme.
 SOURCE Beta vulgaris.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Caryophyllaceae; Chenopodiaceae; Beta.

REFERENCE 1
 Kloos,D.U., Oltmanns,H., Dock,C., Stahl,D. and Hehl,R.
 Isolation and molecular analysis of six taproot expressed genes from sugar beet
 J. Exp. Bot. 53 (373), 1533-1534 (2002)
 MEDLINE 22015900
 PUBMED 12021303
 REFERENCE 2 (bases 1 to 1980)
 TITLE Direct Submission
 AUTHORS Hehl,R.
 JOURNAL Submitted (05-APR-2001) Hehl R., Institute of genetics, Technical University Braunschweig, Spielmannstr. 7, Braunschweig, D-38106, GERMANY

FEATURES
 Source 1..1980
 /organism="Beta vulgaris"
 /db_xref="taxon:161934"
 /country="Germany"
 44..1504
 /codon_start=1
 /product="putative vacuolar processing enzyme"
 /protein_id="CA43295.1"
 /db_xref="GI:14594818"
 /translation="MMIRYTSGLVILVLCVLMSSVDSRLMVDNLIRWPSDHPISFESDDSGVTWAVLIAGSSGYMNYRQADYCHAYQVKKGLKQDENLIVFTDDJAYDEENPRCVLINSYGHVDYAPKPYTGEDVTNNFFAALGNKDAITCGSKVNSGPNHIFPVLSDHGGAVLGNTPYLYADELIETLKEKASGYTSLVYIEACSGSIFEGILPEGLNIVATTASNAVSSWYTCGQDPNPVEYDTCGLDGLYSVNIEDSERHNLHTSELKOQYEVVKTAKETAEKFGSHVNYQDKELTQDMLYLYMGTPNNNNTYVDDNSLHTSSNAVQNRDADLIHFNPKFKASEGSQRKINAQOFMEVMSHRVHLDSDIKLIKLLFGLEKGLVLTQVPTGQPLVDWMLCLATLTFTFKIKICGSLSQYGMKIMRSTIANICNAGITTNMAEASAOACFPSPSGSLHRGFS"

BASE COUNT 558 a 369 c 455 g 598 t
 ORIGIN

Alignment Scores:

Pred. No.: 4,13e-100 Length: 1980
 Score: 1236.00 Matches: 232
 Percent Similarity: 66.81% Conservative: 78
 Best Local Similarity: 50.00% Mismatches: 142
 Query Match: 50.37% Indels: 12
 DB: 8 Gaps: 2

US-09-934-066-2 (1-466) x BVU309173 (1-1980)

QY	10	IleLeuValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThrGlnLeu	29
Db	68	GTACTGATAGTGTATGCGTTTGTGATGAGCTCGGTGCTGATAGCCGT-----CTG	118
QY	30	LeuAsnAspAsn-----AspValGluSerSerAspLys	40
Db	119	ATGGTGGATAACCTCATCAGATGGCGCTCTGATCATCTTCTATCTTTGAGAGCGATGAT	178
QY	41	SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr	60
Db	179	GATTCAGTAGGACTAGATGGCGGTCTGATTCGCGGATCTAGTGGCTACTGGAATTAC	238
QY	61	ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp	80
Db	239	AGGCATCAGCAGATGTTGTCTATGCTTACCAAGTCTAAAAAGAGGCTCTAAAGGAT	298
QY	81	GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro	100
Db	299	GAACAACATCATCTTTTCATGTATGATGACATTTGCTTATCAGCAAGAGATCCAAACCT	358
QY	101	GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr	120
Db	359	GGAGTTCTCATTAATAGTCTTATGGACATGATGTTTATCGAGGAGTACCAAGAGTAT	418
QY	121	ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly	140
Db	419	ACAGGGAGATGCTGCTCAATAACTTCTTCGACGCTATCTTGGAAACAAAGATGCC	478
QY	141	ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr	160
Db	479	ATTACTGGTGGCAGCGAAGGTGGTGAATAGCGGACCTAATGATCAGATCTTCATCTTT	538
QY	161	TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla	180
Db	539	TACTCAGACCATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	598
QY	181	LysAspPheAsnGluValLeuLysMetHisLysArgLysLysTyrAsnLysMetVal	200
Db	599	GATGAGCTCATTTGAAGGAAACATGCGCTCTGGAACCTTATAAAGCCCTGGTA	658
QY	201	IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu	220
Db	659	GTCTATATGAAGCCGTGAATCTGGAGCATCTTTGAGGGTATTCTTCTGAAGGACTT	718
QY	221	AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPro	240
Db	719	AATATATATGCTACCACTGCTCAAAATGCTGTAGAGAGTAGTTGGGAACTTATGCTCT	778
QY	241	GluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle	260
Db	779	GGACAGATCTTAATCTTCACAGATGATGATACCTGTTGGGTGATTTGATACACCGG	838
QY	261	SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyr	280
Db	839	TCFTGGATCGAAGTAGTGGAGGCACAACTTCACACGCGAATCTTTGAAGCAGCAATAC	898
QY	281	HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe	300
Db	899	GAAGTGGTTAAGCAGGAGCAGCAGAAAAAGCCCTTTCTACGGCTCATGTCATGTCAGTAT	958
QY	301	GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn	320
Db	959	GGTGATAAAGCTTACCGGAGGACATGCTTTACTTGTACATGGGTACAAATCTTAATAAT	1018

QY	321	AspAsnPhenThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn	340
Db	1019	GAATAATACATATGTTGGATGATAATTCGTTGCATCTTCTTAATGCTGTTAAT	1078
QY	341	ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer	360
Db	1079	CAACGAGATGCCGATCTTATCCATTTCTGGAAACAGTTTTCGCAAGGCTTCTGAAGGATCT	1138
QY	361	LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle	380
Db	1139	CAAGAAGAAATAATCCAGAGAGAGTTTATGGAGGTGATGTCACAGAGTGCACCTTG	1198
QY	381	AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu	400
Db	1199	GATGACAGCATCAAACTCATTTGGAAGCTTCTATTTGGAATCAGAGGCGCTTAGGTGTG	1258
QY	401	LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr	420
Db	1259	CTACAAACTGTTCGACCTACAGACAACCCCTCGTCGATGACTGGAACCTGCCTGAAACA	1318
QY	421	LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr	440
Db	1319	TTGGTAAAGACATTTGAAAGCATTTGTTGATCTCTATCTCAATACGGAATGAAACACATG	1378
QY	441	GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle	460
Db	1379	CGTTCGATGCAAAACATATGCAATGCGGATCACAACATCAATCAATGCGGAGCTTCA	1438
QY	461	GluGlnAlaCys	464
Db	1439	GCACAAGCTTGC	1450
RESULT 15			
AY090296			
LOCUS	AY090296	1437 bp	linear PLN 24-MAR-2002
DEFINITION	Arabidopsis thaliana At2g25940/F17H15.3 mRNA, complete cds.		
ACCESSION	AY090296		
VERSION	AY090296.1	GI:19699181	
KEYWORDS	FLI CDNA		
SOURCE	Arabidopsis thaliana		
ORGANISM	Arabidopsis thaliana		
REFERENCE	1 (bases 1 to 1437)		
AUTHORS	Kim C.J., Chen, H., Cheuk, R., Meyers, M.C., Shinn, P., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.		
TITLE	Arabidopsis ORF clones		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1437)		
AUTHORS	Kim, C.J., Chen, H., Cheuk, R., Meyers, M.C., Shinn, P., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.		
TITLE	Direct submission		
JOURNAL	Submitted (14-MAR-2002) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA		
COMMENT	RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,		

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2003, 15:08:47 : Search time 378 Seconds
(without alignments)
9293.961 Million cell updates/sec

Title: US-09-934-066-1
Perfect score: 1560
Sequence: 1 ctcaagaatcagattca.....gaaaaaaaaaaaaaaa 1560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1518.8	97.4	1541	21 AAC35910	Arabidopsis thalia
2	1401	89.8	1401	21 AAC50984	Arabidopsis thalia
3	454.8	29.2	1485	21 AAC42934	Arabidopsis thalia
4	452.8	28.0	1742	14 AAQ50570	Asparaginylenoep
5	451	28.9	1323	14 AAQ50560	Asparaginylenoep
6	442	28.3	1323	14 AAQ50562	Asparaginylenoep
7	442	28.3	1749	14 AAQ50574	Asparaginylenoep
8	441.6	28.3	1323	14 AAQ50561	Asparaginylenoep
9	441.6	28.3	1830	14 AAQ50572	Asparaginylenoep

10	345.6	22.2	1910	14 AAQ50579	Asparaginylenoep
11	345	22.1	1323	14 AAQ50559	Asparaginylenoep
12	343.4	22.0	1323	14 AAQ50563	Asparaginylenoep
13	323.6	20.7	1640	14 AAQ50575	Asparaginylenoep
14	319.2	20.5	1152	14 AAQ50564	Asparaginylenoep
15	305.4	19.6	894	14 AAQ50565	Asparaginylenoep
16	302.4	19.4	813	14 AAQ50566	Asparaginylenoep
17	281	18.0	753	14 AAQ50569	Asparaginylenoep
18	229.2	14.7	1453	14 AAQ50573	Asparaginylenoep
19	198.8	12.7	1128	19 AAV29038	Open reading frame
20	198.8	12.7	1749	19 AAV29037	Human protein comp
21	198.8	12.7	1936	19 AAV44686	Osteoclast inhibit
22	198.8	12.7	2030	22 AAC68662	Human FHD02 cDNA
23	190.8	12.2	1855	18 AAT66381	Human adrenal gland
24	190.8	12.2	1855	20 AAT69241	Novel cysteine pro
25	190.8	12.2	1855	21 AAZ44095	Human cysteine pro
26	183.4	11.8	1250	14 AAQ50571	Asparaginylenoep
27	143	9.2	2373	23 AAS65199	DNA encoding novel
28	133.6	8.6	1336	24 ABK48384	DNA encoding human
29	75.6	4.8	423	16 AAT35112	Enhanced expressio
30	73.2	4.7	2135	22 AAC83228	DNA sequence from
31	55.8	3.6	1490	21 AAC36486	Arabidopsis thalia
32	50	3.2	1491	21 AAC45045	Arabidopsis thalia
33	49.2	3.2	329	14 AAQ50578	Asparaginylenoep
34	45	2.9	1492	19 AAV38747	Asparaginylenoep
35	45	2.9	1492	19 AAV38748	Complementary stra
36	45	2.9	5219	24 ABL49351	Human polynucleoti
37	44.6	2.9	1664976	19 AAV21209	Methanococcus jann
38	44.2	2.8	620	7 AAN60479	Encodes P. falcipa
39	43.6	2.8	15061	22 AAL04226	Human reproductive
40	43.6	2.8	15061	23 ABK42104	Genomic sequence #
41	43.4	2.8	1405	22 AAC83230	DNA sequence from
42	42	2.7	17137	24 ABL32191	Human immune syste
43	41.4	2.7	1524	21 AAA70107	Plasmodium falcipa
44	41.4	2.7	168575	22 AAH21613	Human hypocrerin r
45	41	2.6	294	22 AAH70082	Human cervical can

ALIGNMENTS

RESULT 1
AAC35910
ID AAC35910 standard; DNA; 1541 BP.
XX
AC AAC35910;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 11859.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0123788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0128845.

protein identification; signal transduction pathway;
hybridisation assay; genetic mapping; gene expression

RESULT 3
AAC42934
ID AAC42934 standard; DNA; 1485 BP.
XX AC AAC42934;
XX AC AAC42934;
DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37409.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;
XX KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 25-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 04-MAY-1999; 99US-0132048.
XX PR 05-MAY-1999; 99US-0132407.
XX PR 06-MAY-1999; 99US-0132484.
XX PR 06-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0135124.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 24-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 07-JUN-1999; 99US-0137528.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138034.
XX PR 10-JUN-1999; 99US-0138340.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139452.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

PR	20-AUG-1999;	99US-0149929.		Qy	321	CTCGAGAACTCTAGGCTGGAGTTATCATATAAACCAGATGGAGAAAGATGTTTATAA	380
PR	23-AUG-1999;	99US-0149902.		Db	321		380
PR	23-AUG-1999;	99US-0149930.		Qy	381	AGGAGTTCCTAAGGACTACACTAAAGAGCTGTAATGTTCAAAACTTCTACAACTGTT	440
PR	26-AUG-1999;	99US-0150566.		Db	381		440
PR	27-AUG-1999;	99US-0151065.		Qy	441	ACTTGGAAATGAAAGTGGCGTCACAGGAGGAAATGGCAAAAGTTGTGAAAAAGTGTCTAA	500
PR	27-AUG-1999;	99US-0151080.		Db	441		500
PR	30-AUG-1999;	99US-0151303.		Qy	501	TGATAATATCTTCATCTATTATCTGACCATGGAGCTCTGGCTTAATACGGATGCCAC	560
PR	31-AUG-1999;	99US-0151438.		Db	501		560
PR	01-SEP-1999;	99US-0151930.		Qy	561	TGCTGATGAAGTTATGCAAAAGATTTCAATGAAGTCTTGGAAGATGATGAAGCAAA	620
PR	07-SEP-1999;	99US-0152363.		Db	561		620
PR	10-SEP-1999;	99US-0153070.		Qy	621	AAATACAAAGATGGTGATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAG	680
PR	13-SEP-1999;	99US-0153758.		Db	621		680
PR	15-SEP-1999;	99US-0154018.		Qy	681	GATTTTAAAGAAAATCTCAACATATACGCAGTCTGCTGCTAAATTTCTAAAGAGCAG	740
PR	16-SEP-1999;	99US-0154039.		Db	681		740
PR	20-SEP-1999;	99US-0154779.		Qy	741	CTGGGAGTTTACTGTCTCAGTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	800
PR	22-SEP-1999;	99US-0155139.		Db	741		800
PR	23-SEP-1999;	99US-0155486.		Qy	801	CGGGATACATTTAGCATCTCTTGGGTTTGAGGACAGTGCCTTCATGACATGAGCAAGA	860
PR	24-SEP-1999;	99US-0155659.		Db	801		860
PR	28-SEP-1999;	99US-0156458.		Qy	861	GACTTTGGAGCAACAATACCACGTTGTAAGAGAGAGTAGGATCTGATGTAC- - -CAGA	917
PR	29-SEP-1999;	99US-0156596.		Db	861		920
PR	04-OCT-1999;	99US-0157117.		Qy	918	GACTTCTCATGTATGCGGTTTCGGAACAGAGAGTCTTAAAGATTATCTTCTCTCTTA	977
PR	05-OCT-1999;	99US-0157753.		Db	921		980
PR	06-OCT-1999;	99US-0157865.		Qy	978	CATTGGAAGAAATCCTGAAACGATACATTTCTACCTTCACGGAATCTTTTCTCACCAAT	1037
PR	07-OCT-1999;	99US-0158029.		Db	981	TATGGGAACAACCCCTGCCAATGACAATTTTACCTTTGCGGATGCGGAAT- - -TCACTAA	1037
PR	08-OCT-1999;	99US-0158232.		Qy	1038	CTCTAATTTCTGGCTTGGTCAATCCGGCGGATATTTCTCTCTCTATACCTCCAGAGAAAT	1097
PR	12-OCT-1999;	99US-0158369.		Db	1038		1097
PR	13-OCT-1999;	99US-0159293.		Qy	1098	TCAAAAAGCTCCAAATGGGATCACTTGAACCAAGAGAGCTCAGAAGAAATTCGTTGACGA	1157
PR	13-OCT-1999;	99US-0159394.		Db	1098	CCGAAGACCAAGAGGTTTCAAGAGAAACAGAGAGCTCAGAAGCAAGTACTTTGANGC	1157
PR	13-OCT-1999;	99US-0159395.		Qy	1158	AAAGAAATCATAGGAACAAATCGATCAGACATTATACAGACATTCTCGCGCTTTTCAGTTAA	1217
PR	14-OCT-1999;	99US-0160741.		Db	1158	CATGCTCTCAGAGACTTCATATTGACAATAGCGTGATACTCTCGGAAATATCTTGTGG	1217
PR	21-OCT-1999;	99US-0160767.		Qy	1218	ACAAACCAATGCTTAAATCTTTAACTTCCCAAGAACAAACAGGACAGGCTCTTTGAGA	1277
PR	21-OCT-1999;	99US-0160770.		Db	1218	CATTTCCAGAGGCTCTGAAGTCTAAACAAAGTACGGTCTCTGCTGGGCAAGCTCTAGTGA	1277
PR	21-OCT-1999;	99US-0160814.		Qy	1278	CGATTGGGATTTGCTTCAAGACTCTAGTTAATAGCTTCAAGAACTACCTGCGGTGCAACGGT	1337
PR	21-OCT-1999;	99US-0160815.		Db	1278	TGACTGGAATGCTTCAAGAAATCAGGTGAGAGCTTTTCGAGAGGACACTGTGGATCGCTGC	1337
PR	22-OCT-1999;	99US-0160980.		Qy	1338	GCATTACGAGATTGAAGTATACAGGAGCGCTTGGCAATATCTGCAATATGGAGTGGATGT	1397
PR	22-OCT-1999;	99US-0160981.		Db	1338	TCAGTAGCGTATCAAGACATAGGCTCTTTTGGCAACATCTGCAATCCAGGATTCAAAT	1397
PR	25-OCT-1999;	99US-0161404.		Qy	1398	GAAGAAACTGTTTTCAGCCCAATGAACAAAGCTTGTTCGA	1435
PR	25-OCT-1999;	99US-0161405.		Db			
PR	25-OCT-1999;	99US-0161406.					
PR	26-OCT-1999;	99US-0161359.					
PR	26-OCT-1999;	99US-0161360.					
PR	28-OCT-1999;	99US-0161920.					
PR	28-OCT-1999;	99US-0161992.					
PR	28-OCT-1999;	99US-0161993.					
PR	29-OCT-1999;	99US-0162142.					

Query Match

Best Local Similarity

Matches 785; Conservative

29.2%; 60.5%;

Score 454.8; DB 21; Length 1485;

Pred. No. 4.5e-107;

0; Mismatches 507; Indels 6; Gaps 2;

Qy	141	TGTTGAATCTAGGACAGAAGTGCAAAGGCACACGATGGGCTGTTTTAGTTGCTGATC	200
Db	141		
Qy	201	AAATGAATATTATACATAGGATCAGCTGACATATGCCAGCGTATCAGATCTCG	260
Db	201		
Qy	261	AAAAGCGGTTTAAAAGATGAAACATCATTTGTTTTATGTATGATGATATCGCGTTTC	320
Db	261		

OS XX Canavalia ensiformis.
 PN JP05276960-A.
 XX PD * 26-OCT-1993.
 XX PF 07-AUG-1992; 92JP-0231602.
 XX PR 07-FEB-1992; 92JP-0056023.
 XX PA (SYUZU) SYUZU T.
 XX DR WPI; 1993-373587/47.
 XX PT New gene for encoding asparaginyl endo-peptidase - comprises 8
 PT specified DNA sequences
 XX PS Claim 1; Page 13; 35pp; Japanese.
 XX CC A gene encoding asparaginyl endopeptidase is claimed.
 CC 8 DNA sequences are given (AAQ50576-66). The enzyme is a protease
 CC derived from a seed of Canavalia ensiformis which selectively
 CC hydrolyses C-terminus amide bond of L-asparagine residue
 CC (see AAR43033 and AAR43041).
 CC The enzyme is useful for protein fragmentation and enzymatic
 CC peptide synthesis.
 CC The primers given in AAQ50576-68, AAQ50577-77 and AAQ50583-90 were used
 CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
 CC by PCR.
 XX SQ Sequence 1323 BP; 390 A; 244 C; 312 G; 377 T; 0 other;

Query Match 28.3%; Score 441.6; DB 14; Length 1323;
 Best Local Similarity 60.9%; Pred. No. 1.le-103;
 Matches 771; Conservative 0; Mismatches 489; Indels 6; Gaps 3;

169 GGCACACGATGGCGCTTTTACGTCTGGATCAATGAATATATATTAACATACAGCATCAG 228
 7 GGTACACAGATGGCCATCTATCCGCGCTCCATGGCTACTGGAATACAGGCATCAG 66
 229 GCTGACATATCCACCGCTGATCAGATCTCCGAAAGCGGTTTAAAGATCAAAACATC 288
 67 GCTGATATTTGTCGCTATCAATATGAGGAAGGTGGCTGAAAGAGAAATAT 126
 289 ATTGCTTTATGATCATATATCGGTTTCTCGGAGAAATCTAGCGCTGGAGTTATC 348
 127 ATTGTTTATGATGATGACATGCTTCAATGAGAACACCCAAACCTGGAGTCATC 186
 349 ATTAATAACCCAGATGGAGAAGATGTTTATAAGGAGTCTTAAGGACTACACTAAAGAA 408
 187 ATTAACAACCCAGATGGGATGATGTTTATGAAGGAGTTCACAAAGGATACACTGGCGAC 246
 409 GCTGTTAATGTCACAACTTACAAATGTTTACCTTGGAAATGAAAGTGGCGTCACAGGA 468
 247 GATGTTTACTGCTGATCACTTCTATCTGTTATCTTGGAAATGAAATGAGCTTACAGGT 306
 469 GGAATGGCAAAAGTTGTGAAAGAGTGGTCTTAATGATAATATCTTCACTATTATGCTGAC 528
 307 GGCAGTGGGAAGGTTGTGAACAGATGGCGCTGATGATGATGATGATGATGATGATGAT 366
 529 CATGAGCTCTCGGCTTAAATAGCGATGGCCACTGGTGGATGATGATGATGATGATGATGAT 588
 367 CATGAGGTCACAGGAGTCTTGGGATGGCTGGTCTTCTTATATGATGATGATGATGATGAT 426
 589 AATGAAGTCTTGGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 648
 427 ATGAAGTCTTGAAGAAAAAATCACTGCTGTAACGCTATAAAAGCCCTGTTATTTATCTT 486
 649 GAAGCATGTGAATCAGGAAGATGTTTGAAGGAGTATTTAAAGAAAAATCTCAACATATAC 708
 487 GAGGCATGTGAATCTGGGAGTATCTTGAAGGCCCTTCTCTCTGAGATATCAATATTTAT 546
 709 GCAGTGACTGCTGCTAAATTTCTTAAGAGAGACGCTGGGAGTTTACTGCTGAGTCATAT 768

547 GCAACAACAGCTTCCAAATGCAGAAGAACAGCTGGGGAACATATTGCCAGCGAGGAT 606
 769 CTTCTCCCTCCCTTCTGAGATTTGGAACCTTCTCTCGCGGATACATTTAGCATCTCTTGGCTT 828
 507 CCCAGTCTCCCTCCAGATATTTCACTTCTGCTGGTGACGAGTACAGTCTCTTGGATG 566
 825 GAGGACAGTACCTTTCATGACATGAGCAAAAGAGCTTTGGAGCAACATATACCACTTGG-T 887
 667 GAAGACAGTGGAGAGAAATTTCCGACAGCAAACTTTCCACCAACATATGAAATTTGGTT 726
 888 AAAGAGAAGATAGGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 945
 727 AAAGAGAGGACTATTAACGGAAGTATATACCATAGCTCTCAGCTGATGATGATGATGATGAT 786
 946 GAGAAGATGCTTAAAGATTTATCTTCTCTTCACTTTGGAAGAAATCTCTGAAACCAATAC 1005
 787 ATAAGTCTCAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 846
 1006 TTCACCTTTCACGGAATCTTTTCTCTCCCACTCTCTTAATTTCTGGCTTGGTCAATCCGGC 1065
 847 TTTACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 903
 1066 GATATCTCTCTGCTATACCTTCCAGAGAAAGATTTCAAAAGCTTCCAAATGGGATCACTTGA 1125
 904 GTTGTCTGATCTCATCTCATCTTTTGGGAGAAAGTTCCGCAAAAGCTCTGAGGGTTCTTACCAG 963
 1126 AGCAAAAGAGCTCAGAAAGAAATTTGCTTGGAGAAAGAAATCATAGGAAACAAATTCGATCAG 1185
 964 AAAGATGCTCTCAGAAACAAATTTCTGGAAGTAAATGCTCTCAGAAATGCAATATACACAC 1023
 1186 AGCATTACAGACATCTCGCGCTTTCAGTTAAACAAACAAATGCTTAACTCTTTAACT 1245
 1024 AGTCTGAAATTTATGGGAGTCTTTTATTTGGCATTTGAAAGGGTCCAGAACTACTCAAT 1083
 1246 TCCACAGAACACACAGCAGCCTCTTTGAGAGATTTGGGATTTGCTTCAAGACTCTAGTT 1305
 1084 CTGTTAGACCGCTGGAATGCACTTTGTTGATGACTGGGACTGCTGAAATATGTTA 1143
 1306 AATAGCTTCAAGANTCACTCGCGTCCAGCGTGCATTCAGGATTTGAAGTATACAGAGCG 1365
 1144 AGCACTTTTGGAGACATATTTGTTGATGCTCTCTGCTCAGTATGGGATGAAATATATAGGTC 1203
 1366 CTGCGCAATATCTGCAATATGGGAGTGGATGTTGAAGCAAACTGTTTCAAGCCATTTGAACAA 1425
 1204 TTTGCAACATCTGCAATGCAAGATTTAAGAAATGACCAAAATGGCTGATGCTCAGCACAA 1263
 1426 GCTTGT 1431
 1264 GCTTGT 1269

RESULT 9
 AAQ50572
 ID AAQ50572 standard; cDNA to mRNA; 1830 BP.
 XX AC AAQ50572;
 XX DT 24-MAY-1994 (first entry)
 XX DE Asparaginyl endopeptidase clone 103.
 XX KW Asparaginyl endopeptidase; Canavalia ensiformis; seed;
 KW L-asparagine; primer; PCR; protein fragmentation;
 KW peptide synthesis; ss.
 XX OS Canavalia ensiformis.
 XX PH Key Location/Qualifiers
 FT CDS 86..1408
 FT /*tag- a
 XX PN JP05276960-A.

specified DNA sequences

PS Claim 1; Page 12; 35pp; Japanese.

XX A gene encoding asparaginylendopeptidase is claimed.

CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease

CC derived from a seed of Canavalia ensiformis which selectively

CC hydrolyses C-terminus amide bond of L-asparagine residue

CC (see AAR43033 and AAR43041).

CC The enzyme is useful for protein fragmentation and enzymatic

CC peptide synthesis.

CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used

CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,

CC by PCR.

XX Sequence 1323 BP; 385 A; 246 C; 346 G; 346 T; 0 other;

Query Match 22.1%; Score 345; DB 14; Length 1323;

Best Local Similarity 56.3%; Pred. No. 8.6e-79;

Matches 692; Conservative 0; Mismatches 525; Indels 12; Gaps 2;

QY 169 GGCACACGATGGCGTCTTTAGTTCGTGATCAAAATGAATATTATAACTACAGGCATCAG 228

DB 7 GGAACCCGGTGGCGGCTCTCGTCTGCTTCANACGGTTATGGAATATATAGGCATCA 66

QY 229 GCTGACATATGCCACGCGTATCAGATACCTCCGAAAAGCGGTTTAAAAGATGAAAACATC 288

DB 67 GCGGATGTGTCCATGCTTACCAGTGTCTGATATAAAGGTGGAGTCAAAAGAAACATTT 126

QY 289 ATTGTGTTTATGATGATCATATCCGTTTCTCGGAGAACTCTAGCCCTGGAGTTATC 348

DB 127 GTGGTGTGTTATGATGATGATATGCTTATAACGCCATGAATCCGACACCCGGGATCATC 186

QY 349 ATTAATAAACACGATGGAGAGATGTTTATAAAGGAGTCTCTAAGGACTACACTAAAGAA 408

DB 187 ATCAACCATCTCCAGGGCCAGACGTGTATGCTGTGTACCTAAGATTAACCCGGTCAG 246

QY 409 GCTGTTAATGTTCAAACTTCTACAAATGTGTACTTGGAAATGAAAGTGGCGTCAACAGA 468

DB 247 GACGTAACACCTCGAGAACCTATATCTGTCTATCTTGGGGCAAGAGTAAAGTTAAAGT 306

QY 469 GAAATGCAAAAGTTGTGAAAGTGGTCTCTAATGATATAATCTTCTATATATCTGCTAC 528

DB 307 GGAAGTGGCAAGGTGATCAACAGCAATCCGGAGGATAGGATATTATATCTTACTCCGAT 366

QY 529 CATGAGCTCTCGCTTAATACGATGCCACTGTGTGATGAAGTTATGGCAAAAGATTTC 588

DB 367 CATGAGGTCCGGAGTCTTGGGATGCCAACGCCACCATTCGTTTATGCCATGATTTT 426

QY 589 AATGAAGTCTTGGAGAAGATGATGAAGAGAAAAAATACAAAGATGGTGTATCTATGTT 648

DB 427 ATTGATGTTTGAAGAANAACATGCAAGTGGAGGCTACAAGGAGATGGTTATATACATA 486

QY 649 GAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTCAACATATAC 708

DB 487 GAAGCTTGTGAGGTGGGAGCATATTTGAGGGTATTAATGCCCAAGGATCTGAATATTTAT 546

QY 709 CCACTGACTGTCTGCTTAATCTTAAAGAGACAGCTGGGGAGTTTACTGTCTCGAGTCATAT 768

DB 547 GTGCAACTGCTCAATGCCAAGACAGACATTTTGGAACTTATTTCTCTGGGATGAAT 606

QY 769 CCTCTCTCTCTTCTGATTTGGAATTTGCTCTGCGGATACATTTAGCATCTCTTTGGGTT 828

DB 607 CCTCTCCACCAAGAGATGACGTAACCTTGCCTGGGGATTTTATACAGCGTTTCTTTGGATG 666

QY 829 GAGGACAGTGCCTTCATGATGAGCAAGAGACTTTGGAGCAACAATACCACCTTGTAT 888

DB 667 GAAGATGTGAGACTCACAATCTAAAAAGGGAAACGGTACACACCAATACAGTCGCGTA 726

QY 889 AAGAGAAGAGTAGGATCTGATGTACACAGAGC-----TTCATGTTATGCCGTTTCGGA 942

DB 727 AGGAACCGACTTCAAAATCTAACAGCTATAGGTTTGGTTCTCATGTGATGCAATACGGT 786

QY 943 ACAGAGAAGATGCTTTAAAGATATATCTTCTCTTACATTTGGAAGAAATCCTGAAACGAT 1002

DB 787 GACACTAACTATTACTGCTGAAAAGCTTTTACTTGTACCATGGTTTGTATCCTGCCACCGTG 846

QY 1003 AACTTCACITTTACCGGAATCCTTTTCTCCACCAATCTCTAATTTCTGGCTTGGTCAATCGG 1062

DB 847 AACTTTCTCTCC-----ACACAACGGCAACCTAGAAAGCTAAATGGAAGTTGTTAACCCAG 900

QY 1063 CGCGATATTCTCTCTGCTATACCTCCAGAGAAGATTTCAAAAAGCTCCAAATGGGATCATT 1122

DB 901 AGAGATGCGAGAGCTTTTCTCATGTGGCAATGTATCAGAGATCAAAACCATCAACCGGAA 960

QY 1123 GAAACCAAGAAGCTCAGAGAATAATGCTTTGACGAAAAGAAATCATATAGGAAACAAATCGAT 1182

DB 961 AAGAAGACTACATCTCGAAGACAGATTACAGAGACAGTGAAGCATAGGAATCATTTGGAT 1020

QY 1183 CAGAGCATTACAGACATTTCTCGCGCTTTCAGTTAAACAACCAATGTCTTAATCTCTTA 1242

DB 1021 GCAGTGTGGAATGATTGGAGTTTGTGTATGAGCAGGAAAAAGTTCTTCGGTCTCTA 1080

QY 1243 ACTTCCACAAGAACACAGGACAGCCTCTTCTGTAGACGATTGGGATTGGCTTCAAGACTCTA 1302

DB 1081 CATTCGCTGAGGCTCTCTGGTCTGCCCTAGTTGATGATTGGACATGCTTTGAAATCTATG 1140

QY 1303 GTTAATAGCTTCAAGAATCACTGCGGTGCAACGGTGCATTCAGGATTTGAAGTATACAGA 1362

DB 1141 GTTAGAGTGTTCGAAACTCACTGTGGTCACTGACTCAGTATGGCATGAAACACATGCGG 1200

QY 1363 GCGCTTGCCAATATCTGCAATATGGAGT 1391

DB 1201 GCATTCGCAAGCTTTGCAACAGCGCGT 1229

RESULT 12

AAQ50563

ID AAQ50563 standard; cDNA to mRNA; 1323-BP.

XX AC AAQ50563;

XX DT 24-MAY-1994 (first entry)

XX DE Asparaginylendopeptidase sequence (5).

XX KW Asparaginylendopeptidase; Canavalia ensiformis; seed;

XX KW L-asparagine; primer; PCR; protein fragmentation;

XX KW peptide synthesis; ss.

XX OS Canavalia ensiformis.

XX PN JP05276960-A.

XX PD 26-OCT-1993.

XX PF 07-AUG-1992; 92JP-0231602.

XX PR 07-FEB-1992; 92JP-0056023.

XX PA (SYUZ/) SYUZO T.

XX DR WPI; 1993-373587/47.

XX New gene for encoding asparaginyl endo-peptidase - comprises 8

PT specified DNA sequences

XX Claim 1; Page 14; 35pp; Japanese.

XX A gene encoding asparaginylendopeptidase is claimed.

CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease

CC derived from a seed of Canavalia ensiformis which selectively

CC hydrolyses C-terminus amide bond of L-asparagine residue

CC (see AAR43033 and AAR43041).

CC The enzyme is useful for protein fragmentation and enzymatic

CC peptide synthesis.


```

Db      4  TATGGTAATTTATGGCATCAAGCGGATGTGTGCATCTTACCAGTTGCTGATAAAGGT 63
QY      268  GGTTTAAAGATGAACCAATCATTTGCTTTATGTATGATCATATCGGTTTTCTCGGAG 327
Db      64  GGAGTCAAGAAGAACAATTTGGTGTATGTATGATGATATAGCTTATACGCCATG 123
QY      328  AATCCTAGGCGCTGGAGTTATCATTAATAAACAGATGGAGAAGATGTTTATAAAGGAGTT 387
Db      124  AATCCAGACCGGAGTCATCATCAACCATCTCAGGGGCCAGAGGTGTATGCTGGTGA 183
QY      388  CCTAAGGACTACACTAAGAAGCTGTTAATGTTCAAAACTTCTACAACTGTTACTTGA 447
Db      184  CCTAAGGATTACACCGGTGAGGAGCTAACACCTTGAGAACCTATATGCTGTCTTCTGG 243
QY      448  AATGAAGTGGCTTCACAGCAGGAATGGCAAGTTGTTGAAAGTGTCTCTAATGATAAT 507
Db      244  GACAAGAGTAAAGTTAAAGTGGAGTGGCAAGGTGATCAACAGCAATCCGGAGATAGG 303
QY      508  ATCTTCATCTATTATGCTGACCATGGAGCTCTGGCTTAATAGCGATGCCCACTGTGTAT 567
Db      304  ATATTTATATCTACTCCGATCATGGAGTCCCGGAGTTCTTGGGATGCCAAACGCACCA 363
QY      568  GAAGTTATGCAAAAGATTTCAATGAAGTCTTGGAGAAGATGCATAGAGAAAAAATAC 627
Db      364  TTCGTTTATGCCATGGATTTTATTTGATGTTTGAAGAAGAAACATGCAAGTGGAGGCTAC 423
QY      628  AACAAGATGTGTCTATGTTGAAGCATGTGAATCAGAGATGTTTGAAGGATTTTA 687
Db      424  AAGAGATGGTTATATACATAGAGCTTGTGAGAGTGGGAGCATATTTGAGGGTATTAATG 483
QY      688  AAGAAAATCTCAACATATACGAGTGAAGTCTGCTAATTTCTAAAGAGAGCAGTGGGA 747
Db      484  CCCAAGGATCGAATATTTATGTGACAACCTCGCTCAAAATGCACAAGAGAACATTTTGA 543
QY      748  GTTTACTGTCTGAGTCATATCTCTCHCTCTCTGAGATGGAATCTCTCGGGCAT 807
Db      544  ACTATTGTCCTGGGATGAATCTCTCCACGAGAGTACGTAACCTGCTCGGGGAT 603
QY      808  ACATTTAGCATCTCTGCTGGTGGAGCAGTGAACCTTACATGACATGACCAAGAGACTTG 867
Db      604  TTATACAGCTGTTCTGGATGGAGATAGTGAGCTCACAATCTAAAGAGGAACCGTA 663
QY      868  GAGCAACAATACCACGTTGTAAGAGAGAGTAGGATCTGATGTACAGAGA-----CT 921
Db      664  CAACAGCAATACCAGTCGGTAGGAAGAAACGGACTTCAAAATCTAACAGCTATAGTTTGGT 723
QY      922  TCTCATGTATGCCGTTTCGGAACAGAGAGATGCTTAAAGATATCTTTCTCTTACATT 981
Db      724  TCTCATGTGATGCAATACGGTGAACACTAACATTAATGCTGCTGAAAAGCTTTACTGTACCAT 783
QY      982  GGAAGAAATCCTGAAAACGATAACTTCACTTTACGGAATCCTTTTCTCACCACATCTCT 1041
Db      784  GGTTTGATCCTGCCACCGTGAATTTCTCTCC-----ACACACGGCACTTAGAGCT 837
QY      1042  AATCTGGCTGTGTCATTCGCGCGGATTTCTCTGTGTATCTTACCTCCAGAGAAAGATTCAA 1101
Db      838  AAAATGGAAGTTGTTAACCAAGAGAGATGAGAGCTTTTGTTCATGTGGCAAAATGATCAG 897
QY      1102  AAGCTCCATGGATCACTTGAAGCAAGCAAGCTCAGAGAAATGCTTGAAGCAAAAG 1161
Db      898  AGATCAAAACCATCAACCGGAAAAGAGACTCACATCTCTGGAACAGATATACAGAGACAGTG 957
QY      1162  AATCATAGGAACAAATTCGATCAGAGCATTACAGACATTTCTCGGCTTTCAGTTAAACAA 1221
Db      958  AAGCATAGGAATCATTTGGATGGCAGTGTGGAATTTGATTTGGAGTTTGTGTAGGACCA 1017
QY      1222  ACCAATGCTTAATCTCTTAATCTCCACAAGAACAAACAGACAGACCTCTTTGTAGACGAT 1281
Db      1018  GGAAGAAAGTTCTTCGGGTTCTACATTTCCGTGAGGGCTCTGTGCTGCCCTAGTTGATGAT 1077
QY      1282  TGGGATGCTTCAGACTCTAGTTAATAGCTTCAAGAAATCACTGCGGTGCAACGGTGCAT 1341

```

```

Db      1078  TGGACATGCTTGAATATCTATGGTTAGAGTGTTCGAAACTCAGTGTGGTCACTGACTCAG 1137
QY      1342  TACGGATTGAAGTATATACAGAGCGCTTGCCCAATATCTGCAATATGCGAGT 1391
Db      1138  TATGGCATGAACACATGCGGGCATTCGGCAACGCTTTGCAACACGCGCGT 1187
RESULT 14
AAQ50564
ID   AAQ50564 standard; cdna to mRNA; 1152 bp.
XX   AAQ50564;
AC   AAQ50564;
XX   DT
XX   24-MAY-1994 (first entry)
XX   DE
XX   Asparaginylendopeptidase sequence (6).
XX   KW
XX   Asparaginylendopeptidase; Canavalia ensiformis; seed;
XX   L-asparagine; primer; PCR; protein fragmentation;
XX   peptide synthesis; ss.
XX   OS
XX   Canavalia ensiformis.
XX   PN
XX   JF05276960-A.
XX   PD
XX   26-OCT-1993.
XX   PF
XX   07-AUG-1992; 92JP-0231602.
XX   PR
XX   07-FEB-1992; 92JP-0056023.
XX   (SYUZ/) SYUZO T.
XX   WPI; 1993-373587/47.
XX   PT
XX   New gene for encoding asparaginyl endo-peptidase - comprises 8
XX   specified DNA sequences
XX   PS
XX   Claim 1; Page 14-15; 35pp; Japanese.
XX   CC
XX   A gene encoding asparaginylendopeptidase is claimed.
XX   CC
XX   8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
XX   CC
XX   derived from a seed of Canavalia ensiformis which selectively
XX   CC
XX   hydrolyses C-terminus amide bond of L-asparagine residue
XX   CC
XX   (see AAR43033 and AAR43041).
XX   CC
XX   The enzyme is useful for protein fragmentation and enzymatic
XX   CC
XX   peptide synthesis.
XX   CC
XX   The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
XX   CC
XX   in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
XX   CC
XX   by PCR.
XX   SQ
XX   Sequence 1152 bp; 340 A; 210 C; 297 G; 305 T; 0 other;
Query Match 20.5%; Score 319.2; DB 14; Length 1152;
Best Local Similarity 56.3%; Pred. No. 3.7e-72;
Matches 644; Conservative 0; Mismatches 488; Indels 12; Gaps 2;
QY      169  GGCACAGATGGGCTGTTTAGTTGCTGGATCAAAATGAATATTAATACTACAGGCATCAG 228
Db      7  GGTACCCGGTGGGCGGTTCTGCTGGCTTCAACACGGTTATGGAATATTAGGCATCA 66
QY      229  GCTGACATATGCCACGCGTATCAGATATCTCCGAAAAGCGGTTTAAAGATGAAAACATC 288
Db      67  GCGGATGTGTGCCATGCTTACCAGTTCTGCTATAAAGGTGGAGTCAAGAAGAAACATT 126
QY      289  ATTGTGTTTATGATGATGATATCGCTTTTCTCGGAGATCCTAGCCTGGAGTTATC 348
Db      127  GTGGTGTGTTATGATGATGATATAGCTTATAACCCCATGAATCCAGACCCGGAGTCATC 186
QY      349  ATTAATAAACACAGATGGAGAAGATGTTTATAAAGCAGTTCTTAAGAGCTTACACTAAAGAA 408
Db      187  ATCAACCATCTCAGGGGCCAGACGCTGTATGCTGTGTACCTAAGGATTACACCGGTGAG 245

```

Qy 409 GCTGTTAATGTTCAAACTTCTACAATGTTTACTTGGAAATCAAACTGGCGTCACAGG 468
 Db 247 GACGTAAACACCTGAGAACCTATATGCTGTCTATCTTGGGACAAGAGTAAAGTTAAAGT 306
 Qy 469 GGAATGCGCAAGTCTGAAAGTGGTCTAATGATAATATCTTCAATCTATATGCTGAC 528
 Db 307 GGAAGTGGCAAGTGTACACAACTCCGAGGATAGGATATTTATCTTACTCCGAT 366
 Qy 529 CATGGAGTCTGGCTTAATAGCGATGCCACTGGTGTAGTAAAGTTATGGCAAAAGATTC 588
 Db 367 CATGGAGTCTGGGAGTCTTGGATGGCAACGACCACTTCTTACGCCATGATTTT 426
 Qy 589 AATGAAGTCTGGAGAAGTGTATAGAGAAATAATACAAAGATGGTGTATCTATGTT 648
 Db 427 AATGATGTTTGAAGAAGAACATCAAGTGAGGCTACAAAGGATGTTATATACATA 486
 Qy 649 GAAGCATGTGAATCAGAAAGTATGTTTGAAGGATTTTAAAGAAATACTCAACATATAC 708
 Db 487 GAAGCTGTGAGAGTGGGAGCATATTTGAGGGTATATAATGCCCAAGGATCTGAATTTAT 546
 Qy 709 GCAGTGAAGTCTGCTAATTTCAAGAGACGAGCTGGGAGTTTACTGTCTGAGTCATAT 768
 Db 547 GTGACAACTGGCTCAAAATGCACAAAGAACACTTTTGGAACTTATTTGCTGGGATGAT 606
 Qy 769 CCTCTCTCTCTCTCTGAGATGGAACTTCTGTCGCGGATACATTTAGCATCTCTGGCTT 828
 Db 607 CCTCTCTCTCTCTCTGAGATGGAACTTCTGTCGCGGATACATTTAGCATCTCTGGCTT 828
 Qy 829 GAGGACAGTGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 888
 Db 667 GAAGTAGTGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 726
 Qy 889 AAGAGAAGTGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 942
 Db 727 AGAAGAGTGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 786
 Qy 943 ACAGAGAAGTGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1002
 Db 787 GACACTGAAGTGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 846
 Qy 1003 AACTTCACTTCAAGGATCTCTTCTCAACCAATCTCTAATTTCTGGTGGTCAATCCG 1062
 Db 847 AACTTTTCTT-----CCACACACCGGCAACCTAGAGCTAAATGGAAGTTGTTAAACAG 900
 Qy 1063 CGCGATATCTCTGTATATCTCCAGAGAAAGTTCAAAAGCTCCCAATGGGATCACTT 1122
 Db 901 AGAGATGCGAGAGTCTTGTGTCATGTGGCAAAATGTATCAGAGATCAAAACCATCAACCGGAA 960
 Qy 1123 GAAAGCAAAAGAGCTCAGAAAGAAATGCTTGACGAAAGAAATCATAGGAAACAAATCGAT 1182
 Db 961 AAGAAGACTCACATCTCGAAGACAGATACAGAGACAGTGAAGCATAGCAATCATTTGGAT 1020
 Qy 1183 CAGAGCATACAGACATCTCGGCTTTCAGTTAAACAAACCAATGCTTAAATCTCTTA 1242
 Db 1021 GGCAGTGTGGAATGATGAGTCTTGTGTATGATGACCAAGGAAAGTCTTCGGTCTTA 1080
 Qy 1243 ACTTCCACAAAGAACACAGCAGCTCTTGTAGACGATTTGGGATTTGCTTCAAGACTCTA 1302
 Db 1081 CATTCCTGAGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 Qy 1303 GTTA 1306
 Db 1141 GTTA 1144

RESULT 15
 AAO50565
 ID AAO50565 standard; cDNA to mRNA; 894 BP.
 XX
 AC
 XX
 XX
 DT
 XX

DE Asparaginylendopeptidase sequence (7).
 XX Asparaginylendopeptidase; Canavalia ensiformis; seed;
 KW L-asparagine; primer; PCR; protein fragmentation;
 RW peptide synthesis; ss.
 XX Canavalia ensiformis.
 OS
 XX JP05276960-A.
 PN
 XX 26-OCT-1993.
 PD
 XX 07-AUG-1992; 92JP-02311602.
 PF
 XX 07-FEB-1992; 92JP-0056023.
 PR
 XX (SYUZ/) SYUZO T.
 PA
 XX WPI; 1993-373587/47.
 DR
 XX
 XX New gene for encoding asparaginyl endo-peptidase - comprises 8
 PT specified DNA sequences
 PT
 XX
 XX Claim 1; Page 15; 35pp; Japanese.
 PS
 CC A gene encoding asparaginylendopeptidase is claimed.
 CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
 CC derived from a seed of Canavalia ensiformis which selectively
 CC hydrolyses 5'-terminus amide bond of L-asparagine residue
 CC (see ARA3033 and ARA3041).
 CC The enzyme is useful for protein fragmentation and enzymatic
 CC peptide synthesis.
 CC The primers given in AAO50567-68, AAO50576-77 and AAO50583-90 were used
 CC in the isolation of the fragments given in AAO50569-75 and AAO50578-79,
 CC by PCR.
 XX
 SQ Sequence 894 BP; 265 A; 166 C; 229 G; 234 T; 0 other;
 Query Match 19.6%; Score 305.4; DB 14; Length 894;
 Best Local Similarity 60.9%; Pred. No. 1.2e-68;
 Matches 517; Conservative 0; Mismatches 326; Indels 6; Gaps 1;
 Qy 169 GGCACACGATGGCTGCTTTTGTAGTTGCTGATCAATGAATATTAATCACTACAGCATCAG 228
 Db 7 GGTACCCGGTGGCGGCTTCTCGTGGCTTCAACCGTTATGGAATTTATAGGCATCA 66
 Qy 229 GCTGACATATGCCACGCTATCAGATACCTCCAAAAGGCGTTTAAAAGATCAAAACATC 288
 Db 67 GCGGATGTGCGCATCTTACCAGTTGCTGATAAAGGGTGGAGTCAAGAGANAAACAT 126
 Qy 289 ATTGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 348
 Db 127 GTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 186
 Qy 349 ATTAATAAACAGATGGAGAGATGTTTATAAAGGATTCCTAAGGACTACACTAAAGAA 408
 Db 187 ATCAACCATCTCCAGGGGCGACAGCGTATGCTGTGTACCTTAAGGATTAACCGGTGAG 246
 Qy 409 GCTGTTAATTTTCAAACTTCTACAAATGTGTTTACTTGGAAATGAAAGTGGCGTCAAGGA 468
 Db 247 GAGTAAACACCTGAGAACCTATATGCTGTCTTCTTGGGAGCAAGAGTAAAGT 306
 Qy 469 GGAATGGCAAGTGTGTAAGAGTGTCTTAATGATATATCTTCTATCTATCTGAC 528
 Db 307 GGAAGTGGCAAGTGTATCAACAGCAATCCGAGGATAGGATATTTATTTCTACTCCGAT 366
 Qy 529 CATGGAGTCTCTGGCTTAAATAGCGATGCCACTGGTGTAGTAAAGTTATGGCAAAAGATTC 588
 Db 367 CATGGAGTCTCTGGGAGTCTTGGGATGCCAAACGACCATTCGTTTACGCCATGATTTT 426
 Qy 589 AATGAAGTCTTGGAGAAGATGATAGAGAAATAATACAAAGATGTTGATCTATCTGTT 648
 Db 427 ATTGATGTTTGAAGAAGAACATCAAGTGAGGCTACAGGAGATGTTATATACATA 486


```
QY 649 GAAGCATGTGAATCAGGAGTATGTTGAAGGGATTTTAAAGAAAAATCTCAACATATAC 708
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 GAAGCTTGTGAGAGTGGGAGCATATTTGAGGGTATAATGCCCAAGGATCTGAATATTAT 546
QY 709 GCAGTGAAGTGTGCTTAATCTTAAAGAGAGCAGCTGGGGAGTTTACTGCTCCTGAGTCATAT 768
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 GTGACAACGTGCTCAAAATGCAAGAGAACAGTTTGGAACTTATGTCTCTGGGATGAAT 606
QY 769 CCTCTCCCTCTCTGAGATTGGAATCTGTCGCGATACATTTAGCATCTCTTTGGCTT 828
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 CCTCTCCACCAGAGAGTACGTAACTTCCCTGGGGATTTATACAGCGTTTCTTGGATG 666
QY 829 GAGGACAGTGAACCTTCATGACATGAGCAAAAGAGACTTTGGAGCAACAATACCACGTTGTA 888
|| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 GAAGATAGTGAGACTCACAACTCTAAAAAGGGAAACGGTACAACAGCAATACCAGTCGGTA 726
QY 889 AAGAGAAGAGTAGGATCTGATGATACCAGAGA-----CTTCTCATGTATGCCGTTTCGGA 942
| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 727 AGGAAACGGACTTCAAAATCTAACAGCTATAGGTTTGGTTCTCATGTGATGCAATACGGT 786
QY 943 ACAGAGAAGATGCTTAAAGATTATCTTCTCTTACATTGGAAGAAATCCTGAAAACGAT 1002
|| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 787 GACACTAATATTACTGCTGAAAAGCTTACTTGTAACCATGGTTTGTATCCTGCCACCGTG 846
QY 1003 AACTTCACT 1011
||| ||| |||
Db 847 AACTTCTCT 855
```

Search completed: May 29, 2003, 15:24:41
Job time : 402 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2003, 15:17:54 ; Search time 100 Seconds
(without alignments)
4784.157 Million cell updates/sec

Title: US-09-934-066-1
Perfect score: 1560
Sequence: 1 ctcaagaatcagattca.....gaaaaaaaaaaaaaaaaa 1560

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	198.8	12.7	1936	2	US-09-139-424-3
2	198.8	12.7	2030	3	US-08-706-216-3
3	190.8	12.2	1855	2	US-08-928-613-1
4	190.8	12.2	1855	3	US-09-193-524-1
5	52	3.3	7218	1	US-08-232-463-14
6	45	2.9	1492	4	US-08-745-995A-25
7	45	2.9	1492	4	US-08-745-995A-27
8	41.4	2.7	168575	4	US-09-426-290-1
9	40.8	2.6	206	2	US-08-928-613-22
10	40.8	2.6	206	3	US-08-193-524-22
11	40	2.6	859	4	US-09-247-373B-47
12	39.6	2.5	8920	2	US-08-446-855A-1
13	39.6	2.5	8920	4	US-09-150-741-1
14	38.4	2.5	544	4	US-09-280-116-247
15	38.4	2.5	6474	4	US-08-961-527-155
16	38.2	2.4	72928	3	US-09-009-913-1
17	37.8	2.4	155	2	US-08-928-613-5
18	37.8	2.4	155	3	US-09-193-524-5
19	37.4	2.4	5020	4	US-08-961-527-142
20	37.2	2.4	219	2	US-08-928-613-11
21	37.2	2.4	219	3	US-09-193-524-11
22	37.2	2.4	220	2	US-08-928-613-15
23	37.2	2.4	220	3	US-09-193-524-15
24	36.8	2.4	591	3	US-08-882-501-25
25	36.2	2.3	403	4	US-09-385-982-29
26	36.2	2.3	5852	1	US-07-867-106-2
27	36	2.3	1530	4	US-09-134-001C-159

c	28	35.6	2.3	1388	4	US-09-144-776B-9	Sequence 9, Appli
c	29	35.6	2.3	1712	4	US-09-144-776B-5	Sequence 5, Appli
c	30	35.6	2.3	1712	4	US-09-144-776B-7	Sequence 7, Appli
c	31	35.6	2.3	5852	1	US-07-867-106-2	Sequence 2, Appli
c	32	35.2	2.3	929	3	US-08-860-368B-4	Sequence 4, Appli
c	33	35.2	2.3	1158	3	US-08-860-368B-1	Sequence 1, Appli
c	34	35	2.2	4118	1	US-08-119-125A-3	Sequence 3, Appli
c	35	35	2.2	9542	4	US-08-968-685A-9	Sequence 9, Appli
c	36	34.8	2.2	17656	4	US-09-433-579-3	Sequence 24, Appli
c	37	34.4	2.2	591	3	US-08-882-501-24	Sequence 24, Appli
c	38	34.4	2.2	2415	4	US-09-134-001C-2381	Sequence 2381, Ap
c	39	34.2	2.2	1308	4	US-09-150-213-1	Sequence 1, Appli
c	40	34	2.2	199	2	US-08-928-613-18	Sequence 18, Appli
c	41	34	2.2	199	3	US-09-193-524-18	Sequence 18, Appli
c	42	34	2.2	1609	4	US-09-312-285-1	Sequence 1, Appli
c	43	34	2.2	1609	4	US-09-728-764-1	Sequence 1, Appli
c	44	34	2.2	1720	4	US-09-227-357-139	Sequence 139, App
c	45	34	2.2	1972	4	US-09-370-098-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-139-424-3
; Sequence 3, Application US/09139424
; Patent No. 5985832
; GENERAL INFORMATION:
; APPLICANT: Roodman, G. David
; APPLICANT: Reddy, Sakamuri V.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE FOR
; TITLE OF INVENTION: OSTEOCLAST INHIBITORY FACTOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/139,424
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/772,441
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UTSK:295
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1936 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..1434
; US-09-139-424-3

Query Match 12.7% Score 198.8; DB 2; Length 1936;
Best Local Similarity 59.1%; Pred. No. 8.9e-46;

[illegible]

Db	1416	AATAAAGAAAAATTTAATTTATGTATATAATTGTTCTATCTAAATTTATATCTTTAAGAA	1475
Qy	1542	AAAAAAAAAAAAAAAAAAAA	1558
Db	1476	AAAAAAAAAAAAAAAAAAAA	1492

RESULT 7
US-08-745-995A-27/C
; Sequence 27, Application US/08745995A
; Patent No. 6372887

```

APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6372887el Serine Protease Inhibitor
TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
TITLE OF INVENTION: and Uses Thereof
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/745,995A
FILING DATE: 07-NOV-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: PC-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1492 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-745-995A-27

Query Match      2.9%   Score 45;   DB 4;   Length 1492;
Best Local Similarity 51.8%;   pred. No. 0.0058;
Matches 102;   Conservative 0;   Mismatches 95;   Indels 0;   Gaps 0;

Qy 1362 AGCGCTTGCCAATATCTGCAATATGGGAGTCGGATGGAACCAACTGTTTCAGCCATTGA 1421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 AGATATTTTATCATCTACTATTTTCAGTGGTGGATCTTAAGTACAAAATTTATTGTTGAT 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy	1422	ACAAGCTTGTCGATGTAATGATTTGGAAACCAAGTGAATATCGACATTTTAAAAAATACA	148
Db	137	ATATATTTATTTTGTGGAATATTTTTTAAACAAATTTTGCATAAAAAACATAAGACTAAA	78
Qy	1482	AAGTTAATTTCAATAAAACTCGATGTAGAGATGGTTGGTTCATGACTACTACTCTTTACATG	1541
Db	77	AATAAAGAANAATTTAAATTTATGTATAAATTTGTATATACTAAATATATATCTTTAAGAA	18
Qy	1542	AAAAAAAAAAAAAAAAAAAA	1558
Db	17	AAAAAAAAAAAAAAAAAAAA	1
RESULT 8			

```
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Bergling Ran Olafsdottir
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1998-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
; US-09-426-290-1

Query Match      2.7%; Score 41.4; DB 4; Length 168575;
Best Local Similarity 53.4%; Pred. No. 0.57; Indels 0; Gaps 0;
Matches 87; Conservative 0; Mismatches 76;

QY 1397 TGAAGCAACCTGTTTCAGCCATTGAACACCTTGTTCGATGTAATGATTTGCCAAACAAT 1456
DB 148408 TGTATTAAACACTTAGATCATCTTTAAACATATAATATGTTATAAATAATGTCAGTATAAA 148467

QY 1457 GTGATATTCGACTTTAAATAATCAAAAGTTAATTTCAATAAAACTCGATGTAGAGATGCT 1516
DB 148468 TCAATGTCAGCTTAAAGTTAAGACTGTGATTTTAAATAATACTAGATTTAGATAAAA 148527

QY 1517 TGGTTCATGATCACTACTTTTACATGAAAAAATAAATAAATAAATAAATAAATAAATAA 1559
DB 148528 TCAAAATTTGAATGACATTTATTAACCTTAAATAATAAATAAATAAATAAATAAATAA 148570

RESULT 9
US-08-928-613-22
; Sequence 22, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,613
; FILING DATE:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
```

```
US-08-928-613-22
; APPLICATION NUMBER: US/08/928,613
; FILING DATE:
; CLASSIFICATION: 524
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 206 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: Kidney
; CLONE: 195541
; US-08-928-613-22

Query Match      2.6%; Score 40.8; DB 2; Length 206;
Best Local Similarity 59.5%; Pred. No. 0.034;
Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 564 TGATGAAGTTATGCGCAAAAGATTTCATGAAGTCTTGGAGAAGATGCATTAAGAAAAA 623
DB 15 TGAAGATCTTTCATGTAAGGACCTGANTGAGACCCATCCATTACATGTACAAACACAAAAT 74

QY 624 ATACAACAAGATGCTGATCTATGTTGAAGCATGTGAATCAGGAAGTATGTT 674
DB 75 GTACCGAAAGATGCTGTTCTACATTNAGCGCTGTNAGTCTGGGTCCATGTT 125

RESULT 10
US-08-193-524-22
; Sequence 22, Application US/09193524
; Patent No. 6007997
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,613
; FILING DATE:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
```


TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-853-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 206 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: Kidney
CLONE: 195541
US-09-193-524-22

Query Match 2.6%; Score 40.8; DB 3; Length 206;
Best Local Similarity 59.5%; Pred. No. 0.034;
Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 564 TGATGAAGTTATGGCAAAAGATTTCAATGAAGTCTTGGAGAGATGCAATAGAGAAAAA 623
DB 15 TGAAGATCTTCATGTAAAGGACCTGANTGAGACCATTCATTACATGTACAAACACAAAT 74
QY 624 ATACACAGAGTGTGATCTATGTTGAGCATGTCATCAATCAGGAGTATGTT 674
DB 75 GTACCGAAGAGTGTGTTCTACATTNAGGCGCTGTNAGTCTGGGCCATGTT 125

RESULT 11
US-09-247-373B-47
Sequence 47, Application US/09247373B
Patent No. 618954
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1108-A
CURRENT APPLICATION NUMBER: US/09/247,373B
CURRENT FILING DATE: 1999-02-10
PRIOR FILING DATE: 08/924,747
PRIOR FILING DATE: 1997-09-05
NUMBER OF SEQ ID NOS: 56
SOFTWARE: Microsoft Office 97
SEQ ID NO 47
LENGTH: 859
TYPE: DNA
ORGANISM: SOYBEAN
US-09-247-373B-47

Query Match 2.6%; Score 40; DB 4; Length 859;
Best Local Similarity 57.0%; Pred. No. 0.11;
Matches 73; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 527 ACCATGGAGCTCCTGGCTTATAGCGATGCCACTGGTGATGCAAGTATGCGCAAAAGATT 586
DB 322 ACCTTGGGAAGCTGGATGGGTGATGCAATGTCTACCGCGGGGAAGACAGAAAGGCTG 381
QY 587 TCAATGAAGTCTTGGAGAGATGCAATAGAGAAAAATACAAAGATGCTGATCATG 646
DB 382 TGAAGAGCCATAGAAATGATGGAGAAATAGAGAGAGATTAGGGGAAGAAATTTT 441
QY 647 TTGAAGCA 654
DB 442 TTGGAGGA 449

RESULT 12
US-08-446-855A-1/c
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V

APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
TITLE OF INVENTION: phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match 2.5%; Score 39.6; DB 2; Length 8920;
Best Local Similarity 54.9%; Pred. No. 0.44;
Matches 78; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1419 TGAACAGCTTTCGATGTAATGATTTGCAAAACAATGTATATATCCACCTTTAAATA 1478
DB 397 TTAATAAGTTTATATATTTCAAGTAATTTTATAACAATAATGAACACACATATATATA 338
QY 1479 TCAAAAGTTAATTTCAATAAACTCGATGTAGAGATGGTTGGTTTCATGATCTACTTTTAC 1538
DB 337 TATATATATATATATATATATATATATATATATATAAACTTAATGTTTGTATAA 278
QY 1539 ATGAAAAAATAAAAAAAAAAAAA 1560
DB 277 ATATAAGAAAAAATAAAAAAAAA 256

RESULT 13
US-09-150-741-1/c
Sequence 1, Application US/09150741
Patent No. 6183996
GENERAL INFORMATION:
APPLICANT: Stewart et al.
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
Patent No. 6183996
TITLE OF INVENTION: Synthetase II
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/09/150,741
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PL6380
EARLIER FILING DATE: 1992-12-16
EARLIER APPLICATION NUMBER: AU93/00617
EARLIER FILING DATE: 1993-12-02
EARLIER APPLICATION NUMBER: 08/446,855
EARLIER FILING DATE: 1995-07-06
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0

```

; COUNTRY: USA
; ZIP: 70850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILLING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILLING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 155:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-155

Query Match      2.5%; Score 38.4; DB 4; Length 6474;
Best Local Similarity 56.2%; Pred. NO. 0.82;
Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY      1427 CTGTTTCGATGTAAATGATTGGCAAAACAATGTGATATTTCGACTTTTAAAAAATATCAAAGTT 1486
DB      4572 CTTTTTTGAAATATATGTTATTATATCATCTTTTGTTTTCTTAITGTATATATGAAAAC 4513

QY      1487 AATTTCATAAAAACTCGATGTAGAGATGTTGGTTTCATGTACTACTTTTACATCAAAA 1546
DB      4512 ATGTTCAATTAAAAATTTAGAAAAGTGAGAAAACATTTGTGTACTATTATAAAATCAANA 4453

QY      1547 AAAAAAA 1554
DB      4452 TAAAAAA 4445

Search completed: May 29, 2003, 17:10:05
Job time : 212 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2003, 16:32:39 ; Search time 219 Seconds
(without alignments)
9604.793 Million cell updates/sec

Title: US-09-934-066-1
Perfect score: 1560
Sequence: 1 ctcaagaatcagattcaa.....gaaaaaaaaaaaaaaaaa 1560

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 845702 seqs, 674182571 residues
Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1560	100.0	1560	10	US-09-934-066-1
2	832	53.3	4320	10	US-09-934-066-3
3	454.8	29.2	1485	9	US-09-938-842A-1485
4	362.6	23.2	1461	9	US-09-938-842A-927
5	200.4	12.8	1974	9	US-10-208-408-18
6	198.8	12.7	2002	12	US-10-002-600-10
7	192.6	12.3	393	10	US-09-878-374-3641
8	190.8	12.2	1855	9	US-09-967-796-1
9	177	11.3	405	10	US-09-878-574-4049
10	164	10.5	392	10	US-09-878-574-3407
11	114.4	7.3	422	10	US-09-960-352-11028
12	114.4	7.3	426	10	US-09-960-352-1596
13	111.6	7.2	415	10	US-09-960-352-13205
14	102.8	6.6	407	10	US-09-960-352-5110
15	95.2	6.1	415	10	US-09-960-352-10118
16	91.4	5.9	282	10	US-09-878-574-7312
17	89.4	5.7	960	9	US-10-043-487-97
18	79.6	5.1	375	10	US-09-878-574-2361
19	75.8	4.9	455	9	US-09-918-995-8935

20	65.2	4.2	184	10	US-09-960-352-11538	Sequence 11538, A
21	58.4	3.7	367	10	US-09-878-574-1704	Sequence 1704, Ap
22	48.6	3.1	280	10	US-09-878-574-5529	Sequence 5529, Ap
23	47	3.0	273	10	US-09-923-876-4813	Sequence 4813, Ap
24	46.6	3.0	454	10	US-09-878-574-4350	Sequence 4350, Ap
25	43.6	2.8	15061	9	US-10-052-154-991	Sequence 991, App
26	43.6	2.8	15061	9	US-09-764-891-6914	Sequence 6914, App
27	43.6	2.8	15061	10	US-09-764-847-991	Sequence 991, App
28	42.2	2.7	245	10	US-09-960-352-14074	Sequence 14074, A
29	40.8	2.6	206	9	US-09-967-796-22	Sequence 22, Appl
30	39	2.5	1152	9	US-09-984-130-120	Sequence 120, App
31	39	2.5	1833	9	US-09-984-130-117	Sequence 117, App
32	39	2.5	2008	9	US-09-984-130-26	Sequence 26, Appl
33	38.4	2.5	1545	10	US-09-815-242-4399	Sequence 4399, Ap
34	38.4	2.5	1725	10	US-09-815-242-8228	Sequence 8228, Ap
35	38.4	2.5	6948	10	US-09-070-927A-4	Sequence 4, Appli
36	38.2	2.4	250	10	US-09-878-574-7227	Sequence 7227, Ap
37	38.2	2.4	723	10	US-09-770-149-135	Sequence 135, App
38	38.2	2.4	2287	10	US-09-735-169A-7	Sequence 7, Appli
39	38.2	2.4	2287	10	US-09-735-171A-7	Sequence 7, Appli
40	38	2.4	568	10	US-09-864-761-16410	Sequence 16410, A
41	37.8	2.4	155	9	US-09-967-796-5	Sequence 5, Appli
42	37.8	2.4	511	9	US-10-001-887-76	Sequence 76, Appl
43	37.8	2.4	6053	9	US-10-239-676-75	Sequence 75, Appl
44	37.4	2.4	332	10	US-09-964-824A-393	Sequence 393, App
45	37.4	2.4	599	10	US-09-864-761-7949	Sequence 7949, Ap

ALIGNMENTS

RESULT 1
US-09-934-066-1
; Sequence 1, Application US/09934066
; Patent No. US20020108149A1
; GENERAL INFORMATION:
; APPLICANT: Gruis, Darren B.
; APPLICANT: Jung, Rudolf
; TITLE OF INVENTION: Methods of Increasing Polypeptide
; FILE OF INVENTION: Accumulation in Plants
; FILE REFERENCE: 35718/237251
; CURRENT APPLICATION NUMBER: US/09/934.066
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,804
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-934-066-1

Query Match	100.0%	Score 1560;	DB 10;	Length 1560;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1560;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	CTCAAGAATCAGATTCAAGATAGAGTCTTTTCAACAATGTCTAGTCTCTTTGGTCAC	60	
Db	1	CTCAAGAATCAGATTCAAGATAGAGTCTTTTCAACAATGTCTAGTCTCTTTGGTCAC	60	
Qy	61	TTTCAGATCTTGTGTTTCTTCATGCTTGTCTTCTTCAGTGTAGTCCCGGAAAACC	120	
Db	61	TTTCAGATCTTGTGTTTCTTCATGCTTGTCTTCTTCAGTGTAGTCCCGGAAAACC	120	
Qy	121	CAATGCTGACGATAGTGTGTTTCTTCAGTGTAGTGTGTTTCTTCAGTGTAGTGTGTTG	180	
Db	121	CAATGCTGACGATAGTGTGTTTCTTCAGTGTAGTGTGTTTCTTCAGTGTAGTGTGTTG	180	
Qy	181	GCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	240	
Db	181	GCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	240	

QY 385 GTTCTTA----- 391
|||||
Db 2461 GTTCTTAAGGTTCTTATTCTACTCTCTTTTGTGCGTTATTTCTACGTTGAATTCATTTAC 2520
QY 392 -----AGGACTACACTAAGAAGCTGTTA 415
|||||
Db 2521 ATATATATATCAAGTTTGTGTTATTTGGTTAGGCTACACTAAGAAGCTGTTA 2580
QY 416 ATGTTCAAACTTCTCAATAGTGTACTTGGAAATGAAAGTGGCGTCACAGGAGGAAATG 475
|||||
Db 2581 ATGTTCAAACTTCTCAATAGTGTACTTGGAAATGAAAGTGGCGTCACAGGAGGAAATG 2640
QY 476 GCAAAGTGTGAAAGTGTGCTTAATGATAATATCTTCAATCTATTATGCTGACCATGGAG 535
|||||
Db 2641 GCAAAGTGTGAAAGTGTGCTTAATGATAATATCTTCAATCTATTATGCTGACCATGGAG 2700
QY 536 CTCTCGCTTAAT----- 548
|||||
Db 2701 CTCTCGCTTAATAGGTTTCTTAATTTTATGAAATTTATACGTACCATCAATCCATATC 2760
QY 549 -----AGCGATGCCACTG 562
|||||
Db 2761 TATAATAAAGATTTTCTTGATACTACGAAACCGGATTTTCTCAGCGATGCCACTG 2820
QY 563 GTGATGAAGTTATGGCAAAAGATTTTCAATGAAGTCTTGGAGAAGATGCAATAAGAGAAAA 622
|||||
Db 2821 GTGATGAAGTTATGGCAAAAGATTTTCAATGAAGTCTTGGAGAAGATGCAATAAGAGAAAA 2880
QY 623 AATACACAGAT----- 635
|||||
Db 2881 AATACACAGATGGTATATACTCAACCATTCGTTACCTAGCTTTATACATATGTGTTTC 2940
QY 636 -----GGTGATCTATGTTGAAGCATG 656
|||||
Db 2941 TGTTTTGAATCTATGTTGTTGTTTGGATGTTAGTGATCTATGTTGAAGCATG 3000
QY 657 TGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAATCTCAACATATACGAGTGAC 716
|||||
Db 3001 TGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAATCTCAACATATACGAGTGAC 3060
QY 717 TCGTGTAAATCTTAAAGAGAGCAGCTGGGAGTTTACTGTCCTGAGTCATATCCTCTCC 776
|||||
Db 3061 TCGTGTAAATCTTAAAGAGAGCAGCTGGGAGTTTACTGTCCTGAGTCATATCCTCTCC 3120
QY 777 TCCTCTGAGATTGGAATTTCTCGCGGATACATTTAGCATCTCTTGGCTTGA----- 830
|||||
Db 3121 TCCTCTGAGATTGGAATTTCTCGCGGATACATTTAGCATCTCTTGGCTTGAAGCAG 3180
QY 831 ----- 830
Db 3181 GTACTGCAAAACAAAAAGATTCAATCCTTATGGACTATTCGAATGATTTGATTTGTTCTT 3240
QY 831 -----GGACAGTCACTTCATG 847
|||||
Db 3241 GAGAATATTGTTTCATTTGTTCTATGTTTGTGTTGTTGGGACAGTCACTTCATG 3300
QY 848 ACATGAGCAAGAGACTTTGGAGCAACAATACCACGTTGTAAGAGAGAGTAGGATCTG 907
|||||
Db 3301 ACATGAGCAAGAGACTTTGGAGCAACAATACCACGTTGTAAGAGAGAGTAGGATCTG 3360
QY 908 ATGTACAGAGACTTCTCATGTATGCGGTTTTCGGAACAGAGAGATGCTTTAAAGATTATC 967
|||||
Db 3361 ATGTACAGAGACTTCTCATGTATGCGGTTTTCGGAACAGAGAGATGCTTTAAAGATTATC 3420
QY 968 TTTCTCTTACATGGAAGAAATCCTGAAACAGATACCTTTCACGGAATCCTTTT 1027
|||||
Db 3421 TTTCTCTTACATGGAAGAAATCCTGAAACAGATACCTTTCACGGAATCCTTTT 3480
QY 1028 CCTCACCATACTTAATTCGGCTTGGTCAATCCGGCGGATATTCCTCTCTATACCTCC 1087
|||||
Db 3481 CCTCACCATACTTAATTCGGCTTGGTCAATCCGGCGGATATTCCTCTCTATACCTCC 3540

QY 1088 AGAGAA----- 1093
|||||
Db 3541 AGAGAAAGGTGAGCTTTTTCGGGTTTTTTTGATCATTTTAAACGAAAGAGTTTTTCAGCAT 3600
QY 1094 -----AGATTCAAAAGCTCCAATGGGATCACTTTCAAAAGCAA 1130
|||||
Db 3601 GTTTTAAATGTTTAAATCATCTTTAGATTCAAAAGCTCCAATGGGATCACTTTCAAAAGCAA 3660
QY 1131 AGAAGCTCAGAGAAATTTGCTTGACGAAAGAAATCATAGAAACAAATCGATCAGAGCAT 1190
|||||
Db 3661 AGAAGCTCAGAGAAATTTGCTTGACGAAAGAAATCATAGAAACAAATCGATCAGAGCAT 3720
QY 1191 TACAGACATTTCTCGGCTTTTCAGTTAAACAAACCAATGTCTTAAATCTCTTAACTTCCAC 1250
|||||
Db 3721 TACAGACATTTCTCGGCTTTTCAGTTAAACAAACCAATGTCTTAAATCTCTTAACTTCCAC 3780
QY 1251 AGAACAACAGGACAGCCTTCTTGACAGATTCGGATTCGTTCAAGACTCTA----- 1302
|||||
Db 3781 AGAACAACAGGACAGCCTTCTTGACAGATTCGGATTCGTTCAAGACTCTCTAGTAACAA 3840
QY 1303 ----- 1302
Db 3841 CCACATCTCAAACTTTGTTACTTGTCTTACCGCAACCACTTGCATTTATTACTAAACC 3900
QY 1303 -----GTTAATAGCTTCAAGATCACTCGCGTGCAACGGT 1337
Db 3901 AGTGTATATCGAATGAAAAATCGCAGGTTTATAGCTTCAAGATCACTCGCGTGCAACGGT 3960
QY 1338 GCATTACGGATTCAAGTATACAGGAGCGCTTGCCTCAATATCTGCAATATGGAGTGGATGT 1397
|||||
Db 3961 GCATTACGGATTCAAGTATACAGGAGCGCTTGCCTCAATATCTGCAATATGGAGTGGATGT 4020
QY 1398 GAAGCAAACTGTTTTCAGCCATTTGAACAAGCTTCTGATGTAATGATTTTCAAAACAATG 1457
|||||
Db 4021 GAAGCAAACTGTTTTCAGCCATTTGAACAAGCTTCTGATGTAATGATTTTCAAAACAATG 4080
QY 1458 TGATATTCGACTTTAAATAATCAAAAGTTAAATTTCAATAAACTCGATGATAGATGGTT 1517
|||||
Db 4081 TGATATTCGACTTTAAATAATCAAAAGTTAAATTTCAATAAACTCGATGATAGATGGTT 4140
QY 1518 GGTTCATGACTACTTTTACATGAAAA 1545
|||||
Db 4141 GGTTCATGACTACTTTTACATGAAAA 4168

RESULT 3

US-09-938-842A-1485
; Sequence 1485, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1485
; LENGTH: 1485
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1485

Query Match 29.2% Score 454.8 DB 9 Length 1485

Best Local Similarity 60.5%; Pred. No. 8.6e-108;
Matches 785; Conservative 0; Mismatches 507; Indels 6; Gaps 2;

141 TGTGAATCTAGCGAAGAGTCCAAAGGCGACACAGATGGCTGTGTTAGTTGCTGGATC 200
141 TGTGAATCTAGCGAAGAGTCCAAAGGCGACACAGATGGCTGTGTTAGTTGCTGGATC 200
201 AATGAATATTAATCTACAGGCTAGCTGATATGCGGCTATGCGCTGATCAGCTCCG 260
201 TAGCGATATGGAATACAGGCTAGCTGATATGCGGCTATGCGCTGATCAGCTCCG 260
261 AAAGCGGTTTAAAGATGAAACATCATTTGTTTATGATGATGATGATGCGTTTC 320
261 AAAGCGGTTTAAAGATGAAACATCATTTGTTTATGATGATGATGATGCGTTTC 320
261 AAAGCGGTTTAAAGATGAAACATCATTTGTTTATGATGATGATGATGCGTTTC 320
321 CTGGAGATCTTGAAGCTGGAGTTATCATTTAATAAACAGATGAGAGATGTTTATA 380
321 TTACGAGATCTTGAAGCTGGAGTTATCATTTAATAAACAGATGAGAGATGTTTATA 380
381 AGAGTTCTTGAAGCTTACATTAAGAGCTGTTTAAATGTTCAAACTTCTCAATGCTT 440
381 AGAGTTCTTGAAGCTTACATTAAGAGCTGTTTAAATGTTCAAACTTCTCAATGCTT 440
441 ACTTGAATATGAAGTGGGCTCACAGAGGAAATGCAAAAGTTGTGAAAGTGGTCTTAA 500
441 CTTTGAAGATGAAGTGGGCTCACAGAGGAAATGCAAAAGTTGTGAAAGTGGTCTTAA 500
501 TGATAATATCTTCAATCTATTTATGCTGACCATGGAGCTCTGCTGCTTAAATGAGTCC 560
501 TGATAATATCTTCAATCTATTTATGCTGACCATGGAGCTCTGCTGCTTAAATGAGTCC 560
561 TGTGATGAAGTTATGCAAAAGATTTCAATCACTGTTGAGAGATGCAATAGAGAA 620
561 TGTGATGAAGTTATGCAAAAGATTTCAATCACTGTTGAGAGATGCAATAGAGAA 620
621 AATATCAACAGATGCTGATCTATGTTGAAGCTGTTGAAGCTGTTGAAGCTGTTGAAG 680
621 AATATCAACAGATGCTGATCTATGTTGAAGCTGTTGAAGCTGTTGAAGCTGTTGAAG 680
681 GATTTTAAAGAAATCTCAACATATGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 740
681 GATTTTAAAGAAATCTCAACATATGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 740
741 CTGGGAGTTTACTGCTGAGTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 800
741 CTGGGAGTTTACTGCTGAGTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 800
801 CGCGATACATTTAGCATCTCTTGGCTTGGAGGACAGTACCTTCTCATGATGATGATG 860
801 AGGTGACTTGTACAGTGTGCTTGGATGGAAGATAGTGTATGCAATTTACAGACTGA 860
861 GACTTTGGAGCAACATACACCTTCTGTAAGAGAGAGTAGGATCTGATGATGATGATG 917
861 GACTTTGGAGCAACATACACCTTCTGTAAGAGAGAGTAGGATCTGATGATGATGATG 917
918 GACTTTCTCATGTATGCGGTTTGGAGAGAGAGATGCTTAAAGATTTATCTTCTCTCT 977
921 TGGTCTCATGTATGCAATATGCGGATAGGATAGGATAGGATAGGATAGGATAGGAT 980
978 CATTGGAAGAACTTGAAGAGATTAATCTTCAAGGATGCTTCAAGGATGCTTCAAGGAT 1037
981 TATGGGAAACACCTTGGCAATGCAATTTTACCTTGGGATGCGAAT---TCACTTAA 1037
1038 CTCTAATTTGGCTTGGCTCAATCGCGGATATCTCTCTGCTATACCTTCCAGAGAAAT 1097
1038 GGCACCTTTCAAGAGTTTACAAACACCGGATGATGATGATGATGATGATGATGATG 1097
1098 TCAAAAGCTTCAATGGGATCACTTTGAAAGCAAGAGCTCAGAGAAATGCTTGGATG 1157
1098 CCAGAAAGCAGCAAGAGGTTTCAAGAGAAACAGAGAGCTCAGAGAAATGCTTGGATG 1157
1158 AAGAATCATAGGAAACAAATCGATCAGAGCATTTACAGACATCTCTGCGGCTTTCAGT 1217

1158 CATGCTCTACAGACTTCATATTTGACAATAGCGTGATCTCGTCGGAATAATCTTCTTTGG 1217
1218 ACAACCAATGTCTTAAATCTCTTAACTTCCACAAGAACACAGCAGCTCTTCTTTGTA 1277
1218 CATTTCCAGAGGTCCTGAAGTCTAAACAAGTACGGTCTGCTGGCAACCTCTAGTCA 1277
1278 CGATTGGGATGCTTCAAGACTCTAGTTTAAAGTCTTCAAGATCACTCGGCTGCAACGGT 1337
1278 TGACTGGAATGCTTTAAATATCAGGTGAGAGCTTTCCAGAGGCACTGTGGATCGCTGC 1337
1338 GCATTGAGGATGAAAGTATACAGGAGCGCTTCCCAATATCTGCAATATGGAAGTGGATGT 1397
1338 TCAGTACGGTATCAAGCACTAGGCTTTTTCGAACATCTCCAAATGCGAGGATTTCAAT 1397
1398 GAAGCAAACTGTTTTCAGCCATGAAACAGCTTTGTTTCA 1435
1398 GGAGCAAAATGGAGGAGGCGAGCTTCACAGGCTTGTACCA 1435

RESULT 4
US-09-938-842A-927
; Sequence 927, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 927
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-927

Query Match 23.2%; Score 362.6; DB 9; Length 1461;
Best Local Similarity 56.4%; Pred. No. 7.5e-84;
Matches 699; Conservative 0; Mismatches 534; Indels 6; Gaps 1;

154 GACAAGTGCACAGGCGACAGCTGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 213
130 GACGAAGTGGTCTCGTACAGATGGGCGGTTCTCGTCTGTTCTTCTGCTGATATGA 189
214 AACTAGAGGATGAGCTGATATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 273
190 AACTAGAGGATGAGCTGATATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 249
274 AAGATGAAACATCATTTGTTTATGATGATGATGATGATGATGATGATGATGATGATG 333
250 AAGAAAGAAACATGCTGTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 309
334 AGCCCTGGAGTTATCATTAATAAACAGATGAGAGATGTTTATAAGGAGTTTCTTAAAG 393
310 CGTCCGGGACTCTCAACACCATCTGACGGTACGATGTTTACGCGGAGTCCCTTAAG 369
394 GACTACACTAAAGAGCTGTTTAAATGTTTCAAACTTCTCAATGTTTACTTTGGAATGAA 453
370 GACTATCTGTTAGTAGTACGCTTACGGCTGCAAACTTCTACGCTGTACTCTAGGCGAC 429
454 AGTGGCGTCAAGAGGAGAAATGGCAAAAGTTGTGAAAGTGGTCTTAATGATTAATCTTC 513
430 AAGGCTGTTAAAGGTGGAAGCGTAAAGTGCATCGCTAGCAAGCCCAACGATCACATTTTC 489

GENERAL INFORMATION:
APPLICANT: Hopkins, Christopher M.
APPLICANT: Peterson, David P.
APPLICANT: Cocks, Benjamin G.
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
FILE REFERENCE: PA-0042 US
CURRENT APPLICATION NUMBER: US/10/002.600
CURRENT FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: 60/243,521
PRIOR FILING DATE: 2000-10-25
NUMBER OF SEQ ID NOS: 116
SOFTWARE: PERL Program
SEQ ID NO 10
LENGTH: 2002
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Template ID: 1400599.2
US-10-002-600-10

Query Match 12.7% Score 198.8; DB 12; Length 2002;
Best Local Similarity 59.1%; Pred. No. 3e-41;
Matches 429; Conservative 0; Mismatches 267; Indels 30; Gaps 4;
QY 168 AGGCACACGATGGCTGCTTTTAGTTGCTGGATCAAAATGAATATTATACTACAGGCATCA 227
DB 258 AGGCAAGCAGCTGGTGGTGTGCTGGCAGGTTCAAAATGGCTGATATATATAGGCACCA 317
QY 228 GGCTGACATATGCCACGCGTATCAGATACCTCCGAAAGGCGGTTTAAAGATGAAAAACAT 287
DB 318 GGCACACGCTGCCATGCTACCAAGATCAATCCCGCAATGGGATCTCTGACCAACAGAT 377
QY 288 CATTTGTTTATGATGATGATATCGGCTTTTCTCGGAAATCCCTAGGCTGGAGTTAT 347
DB 378 GCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 437
QY 348 CATTATAAAGCAGATGGAGGAGATGTTTATAAGGAGTTCCTTAAGGACTACACTAAGA 407
DB 438 GATCAACAGGCGCAATGGCAGATGCTATACAGGAGTCCCGAAGGACTACACTGGAGA 497
QY 408 AGCTGTTAATGTTCAAACTTCTACAATGTTTACTTGGAAATGAAAGTGGCTGCACAG 467
DB 498 GGATGTTACCCCAACAAATTCCTTGTGTTGAGAGCGGATGCGAAGCAGTGAAGG 557
QY 468 ---AGGAAATGGCAAGTGTGAAAGTGGTCTTAATGATAATATCTTCATCTATTATGC 524
DB 558 CATAGGATCCCGCAAGTCTGAAAGTGGCCCGCAGGATCAGCTGTTTCTTACTTCTAC 617
QY 525 TGACCATGAGCTCCTGGCTTAATAGCGATGCCCACTGGTGTGATGAAGTTATGGCAAGA 584
DB 618 TGACCATGAGTCTACTGGAATACGTTTATGAAAGTGGTGTGATGAAGTCTTCAATGA 674
QY 585 TTTCATGAGCTTTGGAGAGATGCATAAGAGAAAAAATACAAAGATGTTGATGATCA 644
DB 675 CCTGAATGAGACCATTCATACATGTAACAACAAATGTACCGAAGATGGTGTCTTA 734
QY 645 TGTGAGCATGTAATCAGAGATGTTTGAAGGATTTTAAAGAAAAATCTCAACAT 704
DB 735 CATGAGCCCTGAGTCTGGTCCATGATGAA---CCACCTCCCGGATTAACATCAATG 791
QY 705 ATACGAGTACTCTCTTAATCTTAAAGAGACAGCTGGGAGTTTACTGCTGAGTC 764
DB 792 TTATGCACTACTCTGCAACCCAGAGAGTCTGCTAGCCCTGTTACTATGATGAGA- 850
QY 765 ATATCTCTCTCTCTCTGATGATGGAATCTGCTCGGATPACATTTAGCATCTCTG 824
DB 851 -----AGAGTCCAGTACCTGGGGGACTGGTACAGCGTCAACTG 890
QY 825 CTTTGAAGACAGTACTCTCATGATGAGCAAGAGCTTTGGAGCAACATACCACCT 884
DB 891 GATGAGAGCTCGGAGCTGGAGATCTGACTTAAGAGAGCCCTGCAACAGCAGTACCACCT 950

QY 885 TGATAA 890
DB 951 GGATAA 956

RESULT 7
US-09-878-574-3641
Sequence 3641 Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: Thompson, Michael D.
APPLICANT: Li, Rong, Thomas J.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 3641
LENGTH: 393
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: LIB3028-007-Q1-B1-F5
US-09-878-574-3641

Query Match 12.3% Score 192.6; DB 10; Length 393;
Best Local Similarity 68.3%; Pred. No. 5e-40;
Matches 267; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 372 TGTATATAAGGAGTCTCTTAAGGACTACACTAAGAGAGCTGTTAATGTTCAAACTTCTA 431
DB 1 TGTATATAAGGAGTCTCTCAAAAGGATTACACCGCGAAGATGTTTACTGTTGATAACTTTT 60
QY 432 CAATGTGTTTACTTGGAAATGAAAGTGGCTGCACAGAGGAAATGCGCAAAAGTGTGTAAGAA 491
DB 61 TGTCTGTTTACTTGGAAATGAAAGTGGCTGCACAGTCTGCTGGCAGTGGCAAGTGTCTGCACAG 120
QY 492 TGGTCTTAATGATATATCTTCTATCTATGCTGACCATGGAGCTCTGGCTTTAATAGC 551
DB 121 TGGTCTGATGATCATATATTTGATATCTATGCTGACCATGGAGTCTGGGCTGCTCG 180
QY 552 GATCCCACTGGTGTGATGAAGTTATGGCAAAAGATTTCAATGAAGTCTTGAGAGATGCA 611
DB 181 GATCCCTGCTGCTTCTTACTTATACCGCGATGATCTGATTGAAGTCTTGAAAGAAAAAGCA 240
QY 612 TAAGAGAAAAAATACAAAGATGGTGTGATCTATGTTGAAGCATGTGAATCAGGAAGTAT 671
DB 241 TGGTCTGGAACATATAAAACCTAGTATTTTATCTGGAGCATGTGAATCTGGAGTAT 300
QY 672 GTTTGAAGGAGTATTAAGAAAAATCTCAACATATACGAGTACGCTGCTCTAATTTCTAA 731
DB 301 CTTTGAAGGCTCTTCTCTGAAGATATCAATATTTATGAACCACTGCTTCCCAATTCAGA 360
QY 732 AGAGCAGCTGGGAGTCTTACTGCTCTGAG 762
DB 361 AGACAGTACTGGGAACATATTTGCCCGGG 391

RESULT 8
US-09-967-796-1
Sequence 1, Application US/09967796
Patent No. US2002015535A1
GENERAL INFORMATION:
APPLICANT: Diep, Binh
Braxton, Scott M.
Deleane, Angelo M.
TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:

QY	645	TGTTGAAGCAGTGTGAATCAGGAGATATGTTTGAAGGGATTTTAAAGAAATAAATCTCAACAT	704
Db	608	CATTGAGCCCTGTGAGTCTGGGTCCATGATGATGAACCACTGCCGGATAACA---TCAATGT	664
QY	705	ATACCGACTGACTGCTGCTAATTTCTAAAGAGACGAGCTGGGAGTTTACTGTCCCTGAGTC	764
Db	665	TTATGCAACTACTGCTGCGCAACCCAGAGAGTCGTCTACGCTGTTACTATGATGAGA-	723
QY	765	ATATCCTCCTCCTCTCTTGAGATTTGGAATCTGTCTGGCGATACATTTTAGCATCTCTTG	824
Db	724	-----AGAGTCCACGTACCTGGGGGACTGGTACAGCGTCAACTG	763
QY	825	GCTTGAGACAGTGAACCTTCATGACATGAGCAAGAGACTTTTGGAGCAACAATACCAGT	884
Db	764	GATGAAGACTCGGACGTGGAAGATCTGACTAAAGAGACCCCTGCACAAGCAGTACCACCT	823
QY	885	TGTAAA	890
Db	824	GGTAAA	829

RESULT 9
 US-09-878-574-4049
 ; Sequence 4049, Application US/09878574
 ; Patent No. US20020110548A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Byrum, Joseph R.
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Thompson, Michael D.
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated w
 ; TITLE OF INVENTION: Plants
 ; FILE REFERENCE: 38-21(15401)B
 ; CURRENT APPLICATION NUMBER: US/09/878, 574
 ; CURRENT FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: 09/333,535
 ; PRIOR FILING DATE: 1999-06-14
 ; NUMBER OF SEQ ID NOS: 15775
 ; SEQ ID NO 4049
 ; LENGTH: 405
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)...(405)
 ; OTHER INFORMATION: unsure at all n locations
 ; OTHER INFORMATION: Clone ID: LIB3028-002-Q1-B1-C9
 US-09-878-574-4049

Query Match	11.3%	Score 177	DB 10	Length 405
Best Local Similarity	64.9%	Pred. No. 5.7e-36		
Matches 261	Conservative 0	Mismatches 141	Indels 0	Gaps 0
QY	406	GAAGCTGTTAATGTTCAAACCTCTACAATGTGTACTTGGAAATGAAAGTGGCGTCACA	465	
Db				
	2	GAAGATGTTACTGTTGGTAACCTTTTTCGTCGCTTACTTGGAAACAAGTCGGCACTTACT	61	
QY	466	GGAGGAATGGCAAGTGTGTGAAAGGGTCCCTAATGATAATATCTTCATCTAATTAGCT	525	
Db				
	62	GGTGCAGTGGGAAGTTGTGCACAGTGGACCTGATGATCATATATTTGTATACTATACT	121	
QY	526	GACCATGGAGTCTCGGCTTAATACCGCATGCCACTGGTGCATGAGCTATTGGCAAAAGAT	585	
Db				
	122	GACCATGGAGTCTGAGGGGTGCTCGGATGCTGCTGCTCTTATATGCGGATGAT	181	
QY	586	TTCATTAAGTCTTTGGAGAAGATGCATAAGAGAAAAAATACAAAGATGGTGATCTAT	645	
Db				
	182	CTGATTGAAGTCTTGAACAAAAACATGCTTCTGGAACGTATAAAACCTAGTATTTTAT	241	
QY	646	GTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAAATCTCAACATA	705	
Db				
	242	CTCGAGGCATGTGAATCTGGGAGTATCTTTTGAAGGTCTTCTCCCGAAGATATCAATAT	301	

Fri May 30 09:13:12 2003

us-09-934-066-1.rnpb

Page 10

Search completed: May 29, 2003, 18:21:46
Job time : 233 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2003, 15:16:26 ; Search time 2083 seconds
(without alignments)
12129.121 Million cell updates/sec

Title: US-09-934-066-1

Perfect score: 1560

Sequence: 1 ctcaagaatcagattca.....gaaaaaaaaaaaaaaaa 1560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: gb_gss:**

18: em_gss_hum:**

19: em_gss_inv:**

20: em_gss_pln:**

21: em_gss_vrt:**

22: em_gss_fun:**

23: em_gss_mam:**

24: em_gss_mus:**

25: em_gss_other:**

26: em_gss_pro:**

27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
c 1	585	37.5	585	10	AV559121 AV559121
c 2	580.2	37.2	606	10	AV557684 AV557684
c 3	561	36.0	561	10	AV559478 AV559478
c 4	541	34.7	541	10	AV559545 AV559545
c 5	535	34.3	535	10	AV561906 AV561906
c 6	528.6	33.9	535	10	AV562667 AV562667

7	524.2	33.6	538	10	AV555212
c 8	502.8	32.2	506	10	AV557191
c 9	492	31.5	574	10	AV831492
c 10	443	28.4	443	10	AV562308
c 11	437	28.0	437	10	AV563648
c 12	431	27.6	431	10	AV564659
c 13	428.2	27.4	462	10	AV559517
c 14	418	26.8	431	10	AV810306
c 15	411	26.3	411	10	AV819966
c 16	409.4	26.2	427	10	AV560592
c 17	409	26.2	437	10	AV826649
c 18	404.8	25.9	501	10	BE577730
c 19	403.8	25.9	416	10	AV800099
c 20	402.4	25.8	422	10	BE522129
c 21	399	25.6	399	10	AV806388
c 22	385.2	24.7	1981	11	AV104034
c 23	379.8	24.3	391	10	AV565648
c 24	378	24.2	400	10	AV799476
c 25	378	24.2	431	10	AV809370
c 26	367.2	23.5	404	10	AV559374
c 27	363.8	23.3	696	12	BG41524
c 28	354.6	22.7	746	14	BQ871135
c 29	354.6	22.7	760	14	BQ871979
c 30	347.8	22.3	356	10	BE524724
c 31	346.4	22.2	410	10	AV795983
c 32	345.8	22.2	357	10	AV564976
c 33	344.4	22.1	378	10	AV811879
c 34	343.8	22.0	766	12	BG123960
c 35	342	21.9	711	12	BG646037
c 36	340.8	21.8	361	10	AV557646
c 37	340.4	21.8	796	12	BG645154
c 38	332.6	21.3	1651	11	AV110063
c 39	327.2	21.0	349	10	AV560835
c 40	325.8	20.9	776	12	BG581593
c 41	321.4	20.6	655	14	BQ871920
c 42	318.8	20.4	339	10	AV564807
c 43	317.2	20.3	801	13	BI308524
c 44	316	20.3	324	10	AV560421
c 45	307	19.7	420	10	AV560641

ALIGNMENTS

RESULT 1	AV559121/c	AV559121	585 bp	mrna	linear	EST 07-SEP-2000
LOCUS	AV559121/c	AV559121	Arabidopsis thaliana	green silicles	Columbia	Arabidopsis
DEFINITION	AV559121	thaliana	CDNA clone SQ1110F 3', mRNA sequence.			
ACCESSION	AV559121					
VERSION	AV559121.1	GI:8730547				
KEYWORDS	EST					
SOURCE	Chale cress.					
ORGANISM	Arabidopsis thaliana					
REFERENCE	Asakura, E., Nakamura, Y., Sato, S. and Tabata, S.					
AUTHORS	Epikaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
TITLE	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.					
	1 (bases 1 to 585)					
	Asakura, E., Nakamura, Y., Sato, S. and Tabata, S.					
	A large scale analysis of cDNA in Arabidopsis thaliana: Generation					
	of 12,028 non-redundant expressed sequence tags from normalized and					
	size-selected cDNA libraries					
JOURNAL	DNA Res. 7, 175-180 (2000)					
MEDLINE	20363093					
COMMENT	Contact: Erka Asamizu					
	The First Laboratory for Plant Gene Research					
	Kazusa DNA Research Institute					
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan					
	Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/					
FEATURES	Location/Qualifiers					
source	1..585					
	/organism="Arabidopsis thaliana"					
	/strain="Columbia"					

ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 561)
 AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7, 175-180 (2000)
 MEDLINE 20363093
 COMMENT Contact: Erika Asamizu
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 FEATURES
 Location/Qualifiers
 1..561
 Source
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="S118502F"
 /clone_lib="Arabidopsis thaliana green siliques Columbia"
 /tissue_type="green siliques"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 154 a 105 c 113 g 189 t
 ORIGIN
 Query Match 36.0%; Score 561; DB 10; Length 561;
 Best Local Similarity 100.0%; Pred. No. 1.9e-97;
 Matches 561; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 935 GTTCGGACAGAGAAGTGTCTTAAAGATTATCTTCTCTTACATTTGGAAGAATCTG 994
 DB 561 GTTCGGACAGAGAAGTGTCTTAAAGATTATCTTCTCTTACATTTGGAAGAATCTG 502
 QY 995 AAAACGATAACTTCACTTTCACGGAATCTTTTCTCACCATCTCTAATTTCTGGCTTG 1054
 DB 501 AAAACGATAACTTCACTTTCACGGAATCTTTTCTCACCATCTCTAATTTCTGGCTTG 442
 QY 1055 TCAATCCCGGATATCTCTGTATACCTCCAGAGAAATTCAGTTCAGTTCATG 1114
 DB 441 TCAATCCCGGATATCTCTGTATACCTCCAGAGAAATTCAGTTCAGTTCATG 382
 QY 1115 GATCACTTGAAGCAAGAAAGCTCAGAGAAATTCCTTGACGAAAGAAATCATAGGAAC 1174
 DB 381 GATCACTTGAAGCAAGAAAGCTCAGAGAAATTCCTTGACGAAAGAAATCATAGGAAC 322
 QY 1175 AAATCGATCAGACATTCAGACATTCGCGGCTTTCAGTTTAAACAAACCAATGCTTAA 1234
 DB 321 AAATCGATCAGACATTCAGACATTCGCGGCTTTCAGTTTAAACAAACCAATGCTTAA 262
 QY 1235 ATCTCTTAATCCACAGACACAGACAGCCTCTGTAGAGATTTGGATTTGCTTCA 1294
 DB 261 ATCTCTTAATCCACAGACACAGACAGCCTCTGTAGAGATTTGGATTTGCTTCA 202
 QY 1295 AGACTCTAGTTTAAAGCTTCAAGATCACTGCGGTGCAACGGTGCATTACGGATTGAAGT 1354
 DB 201 AGACTCTAGTTTAAAGCTTCAAGATCACTGCGGTGCAACGGTGCATTACGGATTGAAGT 142
 QY 1355 ATACAGGAGCGCTTGCCCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTCA 1414
 DB 141 ATACAGGAGCGCTTGCCCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTCA 82
 QY 1415 CCATTGACACAGCTTGTTCGATGTAATGATTTGCAAACTATTCGACTTTTAA 1474
 DB 81 CCATTGACACAGCTTGTTCGATGTAATGATTTGCAAACTATTCGACTTTTAA 22
 QY 1475 AATATCAAGTTAATTTCAAT 1495
 DB 21 AATATCAAGTTAATTTCAAT 1

RESULT 4
 AV559545
 LOCUS
 DEFINITION Arabidopsis thaliana green siliques Columbia Arabidopsis
 thaliana cDNA clone S119c09f 3', mRNA sequence.
 AV559545
 VERSION AV559545.1 GI:8730971
 KEYWORDS EST.
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 541)
 AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7, 175-180 (2000)
 MEDLINE 20363093
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 FEATURES
 Location/Qualifiers
 1..541
 Source
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="S119c09f"
 /clone_lib="Arabidopsis thaliana green siliques Columbia"
 /tissue_type="green siliques"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 177 a 88 c 128 g 148 t
 ORIGIN
 Query Match 34.7%; Score 541; DB 10; Length 541;
 Best Local Similarity 100.0%; Pred. No. 1e-93;
 Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 324 GGAGAACTCTAGGCTGGAGTTATCATTAATAACAGATGGAGAGATGTTTATAAGG 383
 DB 1 GGAGAACTCTAGGCTGGAGTTATCATTAATAACAGATGGAGAGATGTTTATAAGG 60
 QY 384 AGTTCTTAAGGACTACACTAAAGAGCTGTTAATGTTCAAACTTCTACAATGTTTACT 443
 DB 61 AGTTCTTAAGGACTACACTAAAGAGCTGTTAATGTTCAAACTTCTACAATGTTTACT 120
 QY 444 TGGAAATCAAAAGTGGCGTCACAGGAGAAATGGCAAACTGTTGAAAAGTGGTCTTAATGA 503
 DB 121 TGGAAATCAAAAGTGGCGTCACAGGAGAAATGGCAAACTGTTGAAAAGTGGTCTTAATGA 180
 QY 504 TAATATCTTCATCTATTATGCTGACCATGGAGCTCCTGGCTTAATAGCGATGCCACTGG 563
 DB 181 TAATATCTTCATCTATTATGCTGACCATGGAGCTCCTGGCTTAATAGCGATGCCACTGG 240
 QY 564 TGATCAAGTTATGGCAAAAGATTTCAATGAAGTCTTGGAGAGATGTCATAGAGAAAAA 623
 DB 241 TGATCAAGTTATGGCAAAAGATTTCAATGAAGTCTTGGAGAGATGTCATAGAGAAAAA 300
 QY 624 ATACAACAGATGTTGATCTATGTTGAAGCATGTAATCAGAAAGTATGTTTGAAGGAT 683
 DB 301 ATACAACAGATGTTGATCTATGTTGAAGCATGTAATCAGAAAGTATGTTTGAAGGAT 360
 QY 684 TTTAAAGAAAAATCTCAACATATACGCACTGCTCTAATTTCTAAAGAGAGACAGCTG 743
 DB 361 TTTAAAGAAAAATCTCAACATATACGCACTGCTCTAATTTCTAAAGAGAGACAGCTG 420
 QY 744 GGGAGTTTACTGCTCTGAGTCATATCTCTCCCTCTCTGAGATTTGAACTTTGCTCGG 803

```

Db 421 GGGAGTTTACTGTCTGAGTCAATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGG 480
Qy 804 CGATACATTTAGCATCTCTTGTGCTTGGAGACAGTGCACCTTCATGACATGAGCAAAAGAC 863
Db 481 CGATACATTTAGCATCTCTTGTGCTTGGAGACAGTGCACCTTCATGACATGAGCAAAAGAC 540
Qy 864 T 864
Db 541 T 541

RESULT 5
AV561906/c 535 bp mRNA linear EST 07-SEP-2000
LOCUS AV561906 Arabidopsis thaliana green siliques Columbia Arabidopsis
DEFINITION thaliana cDNA clone S016090F 3', mRNA sequence.
ACCESSION AV561906
VERSION AV561906.1 GI:8733332
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 535)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizukazusa.or.jp. URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
source
Location/Qualifiers
1..535
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="S016090F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 149 a 101 c 108 g 177 t
Query Match 34.3%; Score 535; DB 10; Length 535;
Best Local Similarity 100.0%; Pred. No. 1.5e-92;
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1004 ACTTCACCTTCAGGAATCCCTTCCCTCACCACCTCTCTAATCTGGCTTGGTCAATCCGC 1063
Db 535 ACTTCACCTTCAGGAATCCCTTCCCTCACCACCTCTCTAATCTGGCTTGGTCAATCCGC 476
Qy 1064 GCGATATTCCTCTGTATACCTCCAGAGAAAGATTCAAAAGCTCCCAATGGGATCACTTG 1123
Db 475 GCGATATTCCTCTGTATACCTCCAGAGAAAGATTCAAAAGCTCCCAATGGGATCACTTG 416
Qy 1124 AAAGCAAGAAGCTCAGAGAAATTCCTTGACGAAAAGATTCATAGGAAACAATCGATC 1183
Db 415 AAAGCAAGAAGCTCAGAGAAATTCCTTGACGAAAAGATTCATAGGAAACAATCGATC 356
Qy 1184 AGAGCATTTACAGACATTCCTGGCGCTTTTCAGTTAAACAAACCAATGCTTAAATCTCTTAA 1243
Db 355 AGAGCATTTACAGACATTCCTGGCGCTTTTCAGTTAAACAAACCAATGCTTAAATCTCTTAA 296
Qy 1244 CTTCCACAGAACACAGCAGCCTCTTGTAGACGATTCGGATTCGCTTCAAGACCTAG 1303
Db 295 CTTCCACAGAACACAGCAGCCTCTTGTAGACGATTCGGATTCGCTTCAAGACCTAG 236

```

```

Qy 1304 TTAATAGCTTCAAGAATCACTCGGTGCAACGGTGCATTTACGATTTGAAGTATACAGGAG 1363
Db 235 TTAATAGCTTCAAGAATCACTCGGTGCAACGGTGCATTTACGATTTGAAGTATACAGGAG 176
Qy 1364 CGCTTGGCAATATCTGCAATATATGGAGTGGATGCTGAAGCAAACTGTTTCAGCCATTGAAC 1423
Db 175 CGCTTGGCAATATCTGCAATATATGGAGTGGATGCTGAAGCAAACTGTTTCAGCCATTGAAC 116
Qy 1424 ANGCTTCTTCGATGTAATGATTTGCAAAACATGATATCGACTTTTAAANAATATCAAA 1483
Db 115 ANGCTTCTTCGATGTAATGATTTGCAAAACATGATATCGACTTTTAAANAATATCAAA 56
Qy 1484 GTTAATTTCAATAAACTCGATGATAGATGGTGGTTCATGATACACTTTTAC 1538
Db 55 GTTAATTTCAATAAACTCGATGATAGATGGTGGTTCATGATACACTTTTAC 1

RESULT 6
AV562667
LOCUS AV562667 535 bp mRNA linear EST 07-SEP-2000
DEFINITION AV562667 Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone S0173h03F 3', mRNA sequence.
ACCESSION AV562667
VERSION AV562667.1 GI:8734093
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 535)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizukazusa.or.jp. URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
source
Location/Qualifiers
1..535
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="S0173h03F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 170 a 92 c 114 g 159 t
Query Match 33.9%; Score 528.6; DB 10; Length 535;
Best Local Similarity 99.3%; Pred. NO. 2.4e-91;
Matches 531; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CTCACAGAATTCAGATTCAGAGTAGAAGTTTTCACAACTCTCTAGTCTCTTGGTCCAC 60
Db 1 CTCACAGAATTCAGATTCAGAGTAGAAGTTTTCACAACTCTCTAGTCTCTTGGTCCAC 60
Qy 61 TTTTCAGATTCCTCTTTTCTTCATGCTTTGCTTATCTCTCTCAGCTGAGTCCGCAAAACC 120
Db 61 TTTTCAGATTCCTCTTTTCTTCATGCTTTGCTTATCTCTCTCAGCTGAGTCCGCAAAACC 120
Qy 121 CAATTCGTGAACATATGATGTTGAATCTACGGACAGAGTGAAGGACACAGATGG 180
Db 121 CAATTCGTGAACATATGATGTTGAATCTACGGACAGAGTGAAGGACACAGATGG 180
Qy 181 GCCTGTTTACTGCTCGATCAATGAATATTAATACTACAGGCATCAGCGTGCATATGC 240
Db 181 GCCTGTTTACTGCTCGATCAATGAATATTAATACTACAGGCATCAGCGTGCATATGC 240

```



```

Db      181 GCTGTTTTAGTCTGATCAATGAATATATATAACTACAGGCAATCAGGCTGACATATGC 240
Qy      241 CACGCGTATCAGATACTCCGAAAGGCGGTTTAAAAGATGAAACATCATTTGTGTTATG 300
Db      241 CACGCGTATCAGATACTCCGAAAGGCGGTTTAAAAGATGAAACATCATTTGTGTTATG 300
Qy      301 TATGATGATATCGGTTTCTCGGAGAAATCCTAGGCTGGAGTTATCATTAATAACCA 360
Db      301 TATGATGATATCGGTTTCTCGGAGAAATCCTAGGCTGGAGTTATCATTAATAACCA 360
Qy      361 GATGGAGAAGATGTTTATAAAGAGATTCTTAAGGACTACACTAAAGAGCTGTTAATGTT 420
Db      361 GATGGAGAAGATGTTTATAAAGAGATTCTTAAGGACTACACTAAAGAGCTGTTAATGTT 420
Qy      421 CAAACTTCTACATGTTTACTTGGAAATCAAAGTGGCTCACAGGAGGAAATGGCAAA 480
Db      421 CAAACTTCTACATGTTTACTTGGAAATCAAAGTGGCTCACAGGAGGAAATGGCAAA 480
Qy      481 GTTGTGAAAAGTGGTCTTAATGATAATATCTTCAATCATTAATGCTGACCATGGAG 535
Db      481 GTTGTGAAAAGTGGTCTTAATGATAATATCTTCAATCATTAATGCTGACCATGGAG 535

RESULT 7
AV555212 538 bp mRNA linear EST 06-SEP-2000
LOCUS
DEFINITION Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQ008c10F 3', mRNA sequence.
ACCESSION AV555212
VERSION   AV555212.1 GI:8726627
KEYWORDS  EST.
SOURCE    thale cress.
ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 538)
AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
JOURNAL   DNA Res. 7, 175-180 (2000)
MEDLINE   20363093
COMMENT   Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES             source
            Location/Qualifiers
            1..538
               /organism="Arabidopsis thaliana"
               /strain="Columbia"
               /db_xref="taxon:3702"
               /clone="SQ008c10F"
               /clone_lib="Arabidopsis thaliana green siliques Columbia"
               /tissue_type="green siliques"
               /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
               XhoI"

BASE COUNT  168 a  96 c  116 g  158 t
ORIGIN
Query Match      33.6%; Score 524.2; DB 10; Length 538;
Best Local Similarity 98.5%; Pred. No. 1.7e-90;
Matches 529; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy      4 ACAGATCATGATCAAGATAGAGTTTTCACAAATGCTAGTCTCTTGTCACCTTT 63
Db      2 ACAAGATCATGATCAAGATAGAGTTTTCACAAATGCTAGTCTCTTGTCACCTTT 61
Qy      64 CAGATTCCTGTTTCTTCATGCTTTGCTTATCTTCTCAGCTGAGTCCCGCAAAACCCAA 123
Db      62 CAGATTCCTGCGTCTCTTTCTTCTGCTTATCTTCTCAGCTGAGTCCCGCAAAACCCAA 121

```

```

Qy      124 TTCTGAACGATATATGATGTTGAATCTAGCGACAAGAGTGAAGGACACACATGGGCT 183
Db      122 TTCTGAACGATATATGATGTTGAATCTAGCGACAAGAGTGAAGGACACACATGGGCT 181
Qy      184 GTTTTGTGCTCGATCAATGAATATTAATACTACAGGATCAGGCTGACATATGCCAC 243
Db      182 GTTTTGTGCTCGATCAATGAATATTAATACTACAGGATCAGGCTGACATATGCCAC 241
Qy      244 GCCTATCAGATCTCCGAAAGGCGGTTTAAAAGATGAAACATCATTTGTGTTATGAT 303
Db      242 GCCTATCAGATCTCCGAAAGGCGGTTTAAAAGATGAAACATCATTTGTGTTATGAT 301
Qy      304 GATGATATCGGTTTCTCGGAGAAATCCTAGGCTGGAGTTATCATTAATAACCAT 363
Db      302 GATGATATCGGTTTCTCGGAGAAATCCTAGGCTGGAGTTATCATTAATAACCAT 361
Qy      364 GGAGAAGATGTTTATAAAGAGTTCTTAAGGACTACACTAAAGAGCTGTTAATGTTCAA 423
Db      362 GGAGAAGATGTTTATAAAGAGTTCTTAAGGACTACACTAAAGAGCTGTTAATGTTCAA 421
Qy      424 AACTTCTACATGTTTACTTGGAAATCAAAGTGGCTCACAGGAGGAAATGGCAAGTT 483
Db      422 AACTTCTACATGTTTACTTGGAAATCAAAGTGGCTCACAGGAGGAAATGGCAAGTT 481
Qy      484 GTGAAAGTGGTCTTAATGATAATATCTTCAATCATTAATGCTGACCATGGAGCTCT 540
Db      482 GTGAAAGTGGTCTTAATGATAATATCTTCAATCATTAATGCTGACCATGGAGCTCT 538

RESULT 8
AV557191/c 506 bp mRNA linear EST 07-SEP-2000
LOCUS
DEFINITION Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQ062f03F 3', mRNA sequence.
ACCESSION AV557191
VERSION   AV557191.1 GI:8728606
KEYWORDS  EST.
SOURCE    thale cress.
ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 506)
AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
JOURNAL   DNA Res. 7, 175-180 (2000)
MEDLINE   20363093
COMMENT   Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES             source
            Location/Qualifiers
            1..506
               /organism="Arabidopsis thaliana"
               /strain="Columbia"
               /db_xref="taxon:3702"
               /clone="SQ062f03F"
               /clone_lib="Arabidopsis thaliana green siliques Columbia"
               /tissue_type="green siliques"
               /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
               XhoI"

BASE COUNT  134 a  96 c  105 g  171 t
ORIGIN
Query Match      32.2%; Score 502.8; DB 10; Length 506;
Best Local Similarity 99.6%; Pred. No. 2.1e-86;
Matches 504; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      984 AAGAAATCCTGAAAGATTAACCTTCACTTTCAGGGAATCCTTTCTCACCACATCTCTAA 1043

```

```

Db      506  AAGAAATCTCTGAAACGATAAATCTCACTTTCACGGAATCCTTTCTCTCACCACATCTCTAA 447
Qy      1044 TTCTGGCTTGGTCAATCCGCGGATATTCCTCTGCTATACCTCCAGAGAAAGATTCRAAA 1103
Db      446  TTCTGGCTTGGTCAATCCGCGGATATTCCTCTGCTATACCTCCAGAGAAAGATTCRAAA 387
Qy      1104 AGTCCCAATGGGATCTTGAAGCAAGAAAGCTCAGAGAAATGCTTGGACGAAAGAA 1163
Db      386  AGTCCCAATGGGATCTTGAAGCAAGAAAGCTCAGAGAAATGCTTGGACGAAAGAA 327
Qy      1164 TCATAGGAACAATCGATCAGAGCATACAGACATTCCTGGGCTTTTCAGTTAAACAAC 1223
Db      326  TCATAGGAACAATCGATCAGAGCATACAGACATTCCTGGGCTTTTCAGTTAAACAAC 267
Qy      1224 CAATGCTTAAATCTTAACTTCCACAGAACACAGCAGCCTCTTGTAGACGATTG 1283
Db      266  CAATGCTTAAATCTTAACTTCCACAGAACACAGCAGCCTCTTGTAGACGATTG 207
Qy      1284 GGATGCTTCAAGCTCTAGTTAATAGCTTCAAGAAATCACTGCGTCCCAACGGTGCATTA 1343
Db      206  GGATGCTTCAAGCTCTAGTTAATAGCTTCAAGAAATCACTGCGTCCCAACGGTGCATTA 147
Qy      1344 CGGATGGAATATACAGAGCGCTTGCATATATCTGCAATATGGAGTGGATGGAAGCA 1403
Db      146  CGGATGGAATATACAGAGCGCTTGCATATATCTGCAATATGGAGTGGATGGAAGCA 87
Qy      1404 AACTGTTTCAAGCTTGAACAGCTTGTGATGTAATGATTGCAAAACAATGTGATAT 1463
Db      86  AACTGTTTCAAGCTTGAACAGCTTGTGATGTAATGATTGCAAAACAATGTGATAT 27
Qy      1464 TCGACTTTTAAATATCAAGCTTAAT 1489
Db      26  TCGCTTTTAAATATCAAGCTTAAT 1

RESULT 9
AV831492 LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AV831492 574 bp mRNA linear EST 01-APR-2002
AV831492 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-89-M04 5',
mRNA sequence.
AV831492.1 GI:19873552
EST.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 574)
Seki.M., Narusaka.M., Ishida.J., Kamiya.A., Satou.M., Nakajima.M.,
Oono.Y., Sakurai.T., Carninci.P., Kawai.J., Itoh.M., Ishii.Y.,
Arakawa.T., Shibata.K., Shinagawa.A., Muramatsu.M., Hayashizaki.Y.
and Shinozaki.K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@tc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998) cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda phage-1 vector (Carninci et
al., submitted for publication) digested with BamHI and XhoI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further
details.
Location/Qualifiers
1..574
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

FEATURES
source

```

```

/clone="RAFL09-89-M04"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/notes="Site_1: BamHI; Site_2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) treatments"
BASE COUNT 181 a 102 c 119 g 169 t 3 others
ORIGIN
Query Match 31.5% Score 492; DB 10; Length 574;
Best Local Similarity 97.6% Pred. No. 2.3e-84;
Matches 540; Conservative 0; Mismatches 8; Indels 5; Gaps 4;
Qy      1  CTCACAAAGATCAGATTCAGATAGAGATTTTTCACAAAGATCTCTAGTCTCTTGGTCAC 60
Db      22  CTCACAAAGATCAGATTCAGATAGAGATTTTTCACAAAGATCTCTAGTCTCTTGGTCAC 81
Qy      61  TTTTCAGATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 120
Db      82  TTTTCAGATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 141
Qy      121  CAATTCGTAACGATATGATGTTGAATCTAGCGACAAAGAGTGCACAAAGGACACACGATGG 180
Db      142  CAATTCGTAACGATATGATGTTGAATCTAGCGACAAAGAGTGCACAAAGGACACACGATGG 201
Qy      181  GCTGTTTATGTTGCTGGATCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 240
Db      202  GCTGTTTATGTTGCTGGATCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 261
Qy      241  CAGCGTATCAGATATCTCCGAAAGCGGTTTAAAGATGAAACATCATTTGTTTATG 300
Db      262  CAGCGTATCAGATATCTCCGAAAGCGGTTTAAAGATGAAACATCATTTGTTTATG 321
Qy      301  TATGATGATATGCTGTTTCTCGGAGATCTAGGCTGGAGTTATCATTAATAAACCA 360
Db      322  TATGATGATATGCTGTTTCTCGGAGATCTAGGCTGGAGTTATCATTAATAAACCA 381
Qy      361  GATGAGAGATGTTTATAAGAGGTTCTTAAGGACTACACTAAAGAGCTGTTTATGTT 420
Db      382  GATGAGAGATGTTTATAAGAGGTTCTTAAGGACTACACTAAAGAGCTGTTTATGTT 441
Qy      421  CAAACTTCTCAATGTTTCTCGGAGATCTAGGCTGGAGTTATCATTAATAAACCA 476
Db      442  CAAACTTCTCAATGTTTCTCGGAGATCTAGGCTGGAGTTATCATTAATAAACCA 501
Qy      477  CAAAGTCTGGAAGTCTGCTTAAATATATATCTTATGTTGACCATGGA-G 535
Db      502  CAAAGTCTGGAAGTCTGCTTAAATATATATCTTATGTTGACCATGGA-G 561
Qy      536  CTCCTGCTTAAT 548
Db      562  CTCCTGCTTAAT 574

RESULT 10
AV562308/c LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AV562308 443 bp mRNA linear EST 07-SEP-2000
AV562308 Arabidopsis thaliana green silique Columbia Arabidopsis
thaliana cDNA clone SQ187H10F 3', mRNA sequence.
AV562308.1 GI:8733734
EST.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 443)
Asamizu.E., Nakamura.Y., Sato.S. and Tabata.S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and

```

JOURNAL
MEDLINE
COMMENT

size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

```

FEATURES
source
Location/Qualifiers
1. .443
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="S0167h10f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
115 a 90 c 87 g 151 t
BASE COUNT
ORIGIN

```

Query Match	28.4%	Score 443;	DB 10;	Length 443;
Best Local Similarity	100.0%;	Pred. No. 6.le-75;		
Matches 443;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1045	TCCTGGCTTGTC	CAATCGCGCGGATATTCCTCTGCTATACCTCCAGACAAAGATTCAAAA	1104
Db	443	TCCTGGCTTGTC	CAATCGCGCGGATATTCCTCTGCTATACCTCCAGACAAAGATTCAAAA	384
QY	1105	GCCTCCAAATGGG	ATCACTTTGAAAGCAAAAGAGCTCAGAGAAGAAATTCCTTGACGGAAGAAGAT	1164
Db	383	GCCTCCAAATGGG	ATCACTTTGAAAGCAAAAGAGCTCAGAGAAGAAATTCCTTGACGGAAGAAGAT	324
QY	1165	CATAGGAAACAAAT	TCGATCAGAGCATTTACAGACATTCCTGGCGCTTTCAGTTTAAACAAACC	1224
Db	323	CATAGGAAACAAAT	TCGATCAGAGCATTTACAGACATTCCTGGCGCTTTCAGTTTAAACAAACC	264
QY	1225	AATGCTCTTAAAT	CTCTTAACTTCCACAAGAACAAACAGACAGCCTCTTGTAGACGATTGG	1284
Db	263	AATGCTCTTAAAT	CTCTTAACTTCCACAAGAACAAACAGACAGCCTCTTGTAGACGATTGG	204
QY	1285	GATTGCTTCAAGAC	TCTAGTATATAGCTTCAAGAAATCAGTCGCGTGCACCGTGCAATAC	1344
Db	203	GATTGCTTCAAGAC	TCTAGTATATAGCTTCAAGAAATCAGTCGCGTGCACCGTGCAATAC	144
QY	1345	GGATTGAAGTATAC	AGGAGCGCTTGGCAATATCTGCAATATGGGAGTGGATGTCAAGCAA	1404
Db	143	GGATTGAAGTATAC	AGGAGCGCTTGGCAATATCTGCAATATGGGAGTGGATGTCAAGCAA	84
QY	1405	ACTGTTTCAGCCAT	TGAACAGCTTGTTCGATGTATATGGTTCGAAACAAATGTGATATT	1464
Db	83	ACTGTTTCAGCCAT	TGAACAGCTTGTTCGATGTATATGGTTCGAAACAAATGTGATATT	24
QY	1465	CGACTTTAAAAAT	ATACAAAGTTA	1487
Db	23	CGACTTTAAAAAT	ATACAAAGTTA	1

RESULT 11	AV5633648	Arabidopsis thaliana green silicles Columbia Arabidopsis	437 bp	linear	EST 07-SEP-2000
LOCUS	AV5633648/c	thaliana cDNA clone S0191a03F 3', mRNA sequence.			
DEFINITION	AV5633648	Arabidopsis thaliana green silicles Columbia Arabidopsis			
ACCESSION	AV5633648	thaliana cDNA clone S0191a03F 3', mRNA sequence.			
VERSION	AV5633648.1	GI:8735074			
KEYWORDS	EST.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.				
	1 (bases 1 to 437)				
REFERENCE	Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.				
AUTHORS					

TITLE	A large scale analysis of cDNA in <i>Arabidopsis thaliana</i> : Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL	DNA Res. 7, 175-180 (2000)
MEDLINE	20363093
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp , URL: http://www.kazusa.or.jp/en/plant/ .
FEATURES	Location/Qualifiers 1. .437

```

source
1. 437
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ191a03f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
114 a 86 c 85 g 152 t
BASE COUNT
ORIGIN

```

Query Match	28.0%;	Score 437;	DB 10;	Length 437;
Best Local Similarity	100.0%;	Pred. No. 8.6e-74;		
Matches 437;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1059	TCCGCCGGATATTCCTCTGCTATACCTCCAGAGAAAGATTCAAAAAGCTCCAAATGGGATC	1118	
DB	437	TCCGCCGGATATTCCTCTGCTATACCTCCAGAGAAAGATTCAAAAAGCTCCAAATGGGATC	378	
QY	1119	ACTTTGAAAGCAAAAGAAGCTCAGAAGAAATTCGTTGACGAAAAAGCAATCATAGGAAACAAAT	1178	
DB	377	ACTTGAAGCAAAAGAAGCTCAGAAGAAATTCGTTGACGAAAAAGCAATCATAGGAAACAAAT	318	
QY	1179	CGATCAGAGCATTTACAGACATTCCTCGCGCTTTTCAGTTAAACAAACCAATGCTCTTAAATCT	1238	
DB	317	CGATCAGAGCATTTACAGACATTCCTCGCGCTTTTCAGTTAAACAAACCAATGCTCTTAAATCT	258	
QY	1239	CTTAACCTTCACAGAACAAACAGGACAGGCTCTTGTAGACGATTGGGATTCGTTCAAGAC	1298	
DB	257	CTTAACCTTCACAGAACAAACAGGACAGGCTCTTGTAGACGATTGGGATTCGTTCAAGAC	198	
QY	1299	TCTAGTTAATAGCTTCAAGAAATCAGTCGCGTGCACCGGTGCATTACGCGATTCAAGTATAC	1358	
DB	197	TCTAGTTAATAGCTTCAAGAAATCAGTCGCGTGCACCGGTGCATTACGCGATTCAAGTATAC	138	
QY	1359	AGGAGCGCTTGGCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTTCAGCCAT	1418	
DB	137	AGGAGCGCTTGGCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTTCAGCCAT	78	
QY	1419	TGAACAAGCTGTGTCGATGTAATGATTTGCCAAACAAATCTGATATTCGACTTTTAAAAATA	1478	
DB	77	TGAACAAGCTGTGTCGATGTAATGATTTGCCAAACAAATCTGATATTCGACTTTTAAAAATA	18	
QY	1479	TCAAAGTTAATTTCAAT	1495	
DB	17	TCAAAGTTAATTTCAAT	1	

RESULT 12	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	AV564659		AV564659				
		Arabidopsis thaliana green siliques Columbia Arabidopsis thaliana cDNA clone SQ208b11F 3', mRNA sequence.	AV564659				
			AV564659				
			AV564659.1	GI:8736089			
			EST.				
			thale cress				
			Arabidopsis thaliana				
			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 431)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source

1. 431
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ208b11f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 152 a 60 c 104 g 115 t

ORIGIN

Query Match
Best Local Similarity 27.6%; Score 431; DB 10; Length 431;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GGAGATCTAGGCTGGAGTATCATTAATAAACACAGATGGAGAGATGTTTATAAGG 383
Db 1 GGAGATCTAGGCTGGAGTATCATTAATAAACACAGATGGAGAGATGTTTATAAGG 60

QY 384 AGTCTTAAGGCTACATAAGAGCTGTTAATGTTCAAAAGCTTCTACAATGTTGTTACT 443
Db 61 AGTCTTAAGGCTACATAAGAGCTGTTAATGTTCAAAAGCTTCTACAATGTTGTTACT 120

QY 444 TGGAAATGAAGTGGCTACAGAGGAATGGAAGTTGTGAAGTGGTCCCTAATGA 503
Db 121 TGGAAATGAAGTGGCTACAGAGGAATGGAAGTTGTGAAGTGGTCCCTAATGA 180

QY 504 TAATATCTTATCTATTTATCTGACCATGGAGCTCTGCTTAAAGAGTGGCCACTGG 563
Db 181 TAATATCTTATCTATTTATCTGACCATGGAGCTCTGCTTAAAGAGTGGCCACTGG 240

QY 564 TGATGAATATGCAAAAGATTCAATGAAGTCTTGGAGAGATGCATAAGAGAAAAA 623
Db 241 TGATGAATATGCAAAAGATTCAATGAAGTCTTGGAGAGATGCATAAGAGAAAAA 300

QY 624 ATACAACAAGATGGTATCTATCTTGAAGCATGTGAATCAGGAAGTATGTTGAAGGAT 683
Db 301 ATACAACAAGATGGTATCTATCTTGAAGCATGTGAATCAGGAAGTATGTTGAAGGAT 360

QY 684 TTTAAAGAAAAATCTCAACATATACGAGTCTGCTCTAATTTTAAAGAGACAGCTG 743
Db 361 TTTAAAGAAAAATCTCAACATATACGAGTCTGCTCTAATTTTAAAGAGACAGCTG 420

QY 744 GGGAGTTTACT 754
Db 421 GGGAGTTTACT 431

RESULT 13
AV59517/c
LOCUS
DEFINITION
ACCESION
VERSION
KEYWORDS
SOURCE
ORGANISM

AV59517 462 bp mRNA linear EST 07-SEP-2000
thaliana cDNA clone sq118h05f 3, mRNA sequence.
AV59517.1 GI:8730943
EST.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 462)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source

1. 462
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ118h05f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 125 a 92 c 86 g 159 t

ORIGIN

Query Match
Best Local Similarity 27.4%; Score 428.2; DB 10; Length 462;
Matches 436; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1097 TTCAAAAGCTCAATGGATCCTTGAAGCAAGAGCTCAGAGCAATTCCTGGCTTTGAGTGA 1156
Db 462 TTCAAAAGCTCAATGGATCCTTGAAGCAAGAGCTCAGAGCAATTCCTGGCTTTGAGTGA 403

QY 1157 AAAGATCATAGGAACAAATTCGATCAGACATTCAGACATTCCTGGCTTTGAGTGA 1216
Db 402 AAAGATCATAGGAACAAATTCGATCAGACATTCAGACATTCCTGGCTTTGAGTGA 343

QY 1217 ACAACACCATGCTCTTAATCTCTTAAGTTCACAGAACACAGACAGCTCTTTGTAG 1276
Db 342 ACAACACCATGCTCTTAATCTCTTAAGTTCACAGAACACAGACAGCTCTTTGTAG 283

QY 1277 ACGATTGGGATGCTTCAAGACTCTAGTAACTTCAAGAAATCACTCGGTGCAACGG 1336
Db 282 ACGATTGGGATGCTTCAAGACTCTAGTAACTTCAAGAAATCACTCGGTGCAACGG 223

QY 1337 TGCATTACGGATTAAGTATACAGAGCGCTTGCCTAATATCTGCAATATGGAGTGGATG 1396
Db 222 TGCATTACGGATTAAGTATACAGAGCGCTTGCCTAATATCTGCAATATGGAGTGGATG 163

QY 1397 TCAAGCAACTGTTTCAGCCATTCAGACAGCTTGAACAGCTTGTTCGATGAATGTTGCAAAACAT 1456
Db 162 TCAAGCAACTGTTTCAGCCATTCAGACAGCTTGAACAGCTTGTTCGATGAATGTTGCAAAACAT 103

QY 1457 GTGATATTCGACTTTAAATAATCAAGTTAATTTCAATAAACTCGATGTAGAGATGTT 1516
Db 102 GTGATATTCGACTTTAAATAATCAAGTTAATTTCAATAAACTCGATGTAGAGATGTT 43

QY 1517 TGGTTCATGATCTACTTTTACATGAAA 1545
Db 42 TGGTTCATGATCTACTTTTACATGAAA 14

RESULT 14
AV810306/c
LOCUS
DEFINITION
ACCESION
VERSION
KEYWORDS
SOURCE

AV810306 431 bp mRNA linear EST 29-MAR-2002
Arabidopsis thaliana cDNA clone RAFL09-63-G07 3,
mRNA sequence.
AV810306
EST.
thale cress.

```

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.

TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES             Location/Qualifiers
     source           1..431
     /organism="Arabidopsis thaliana"
     /db_xref="taxon:3702"
     /clone="RAFL09-63-G07"
     /clone_lib="RAFL9"
     /dev_stage="plants at various developmental stages from
     germination to mature seeds"
     /lab_host="DH10B"
     /note="Site 1: BamHI; Site 2: SalI; subjected to
     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
     hr) treatments"
BASE COUNT          117 a 89 c 75 g 150 t
ORIGIN
Query Match          26.8%; Score 418; DB 10; Length 431;
Best Local Similarity 99.8%; Pred. No. 3.7e-70;
Matches 429; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1110 AATGGGATCACTTGAAGCAAGAAGCTCAGAAGAAATTCCTGACGAAAGAAATCATAG 1169
      |||||||
Db 431 AATGGGATCACTTGAAGCAAGAAGCTCAGAAGAAATTCCTGACGAAAGAAATCATAG 372
      |||||||

Qy 1170 GAACAAATCGATCAGACGATTACAGACATTCGCGCTTTCAGTTAAACAACCAATGT 1229
      |||||||
Db 371 GAACAAATCGATCAGACGATTACAGACATTCGCGCTTTCAGTTAAACAACCAATGT 312
      |||||||

Qy 1230 CTTAAATCTCTTAATCTCCACAAGAACAGGACAGCCTCTTGTAGACGATTGGGATTG 1289
      |||||||
Db 311 CTTAAATCTCTTAATCTCCACAAGAACAGGACAGCCTCTTGTAGACGATTGGGATTG 252
      |||||||

Qy 1290 CTTCAAGACTCTAGTTAATAGCTTCAAGATCACTGCGGT-GCAACGGTGCAATTCAGGAT 1348
      |||||||
Db 251 CTTCAAGACTCTAGTTAATAGCTTCAAGATCACTGCGGTGGCAACGGTGCAATTCAGGAT 192
      |||||||

Qy 1349 TGAAGTATACAGGAGCGCTGCCAATATCTCCATATGGGAGTGGATGTGAGCAAACTG 1408
      |||||||
Db 191 TGAAGTATACAGGAGCGCTGCCAATATCTCCATATGGGAGTGGATGTGAGCAAACTG 132
      |||||||

Qy 1409 TTTCCAGCATTGAACAGCTTGTTCGATGTAAATGATTGCAAAACAATGTGATATTCGAC 1468
      |||||||
Db 131 TTTCCAGCATTGAACAGCTTGTTCGATGTAAATGATTGCAAAACAATGTGATATTCGAC 72
      |||||||

Qy 1469 TTTAAAATATCAAGTTAATTTCAATAAACTCGATGTAGAGATGGTTGGTTTCATGATA 1528
      |||||||
Db 71 TTTAAAATATCAAGTTAATTTCAATAAACTCGATGTAGAGATGGTTGGTTTCATGATA 12
      |||||||

Qy 1529 CTACTTTTAC 1538

```

```

Db 11 CTACTTTTAC 2
      |||||||

RESULT 15
AV819966/c
LOCUS AV819966 RAFL11 Arabidopsis thaliana cDNA clone RAFL11-08-D05 3',
DEFINITION mRNA sequence.
ACCESSION AV819966
VERSION AV819966.1 GI:19861922
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 411)
REFERENCE
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES             Location/Qualifiers
     source           1..411
     /organism="Arabidopsis thaliana"
     /db_xref="taxon:3702"
     /clone="RAFL11-08-D05"
     /clone_lib="RAFL11"
     /dev_stage="plants at various developmental stages from
     germination to mature seeds"
     /lab_host="DH10B"
     /note="Site 1: BamHI; Site 2: SalI; subjected to various
     treatments (dehydration, cold, high salt, ABA, heat and UV
     ). Dark-grown plants"
BASE COUNT          105 a 82 c 77 g 147 t
ORIGIN
Query Match          26.3%; Score 411; DB 10; Length 411;
Best Local Similarity 100.0%; Pred. No. 8.2e-69;
Matches 411; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1082 ACCTCCAGAGAAAGATTCAAAAAGCTCCAAATGGGATCCTTGAAGCAAGAAAGCTCAGA 1141
      |||||||
Db 411 ACCTCCAGAGAAAGATTCAAAAAGCTCCAAATGGGATCCTTGAAGCAAGAAAGCTCAGA 352
      |||||||

Qy 1142 AGAAATTCCTTGACGAAAGAAATCATAGGAACAAATCGATCAGAGCATTTACAGACATTC 1201
      |||||||
Db 351 AGAAATTCCTTGACGAAAGAAATCATAGGAACAAATCGATCAGAGCATTTACAGACATTC 292
      |||||||

Qy 1202 TCGCGCTTTTCAGTTAAACAACCAATGTCTTAATCTCTTAACCTCCACAGACAACAG 1261
      |||||||
Db 291 TCGCGCTTTTCAGTTAAACAACCAATGTCTTAATCTCTTAACCTCCACAGACAACAG 232
      |||||||

Qy 1262 GACAGCCTCTTGTAGACGATTGGGATTGGCTTCAAGACTCTAGTTAATAGCTTCAAGATC 1321
      |||||||
Db 231 GACAGCCTCTTGTAGACGATTGGGATTGGCTTCAAGACTCTAGTTAATAGCTTCAAGATC 172
      |||||||

```

Qy 1322 ACTGCGGTGCAACGGTGCATTACGGATTGAAGTATACAGGAGCGCTTGCCAAATATCTGCA 1381
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
171 ACTGCGGTGCAACGGTGCATTACGGATTGAAGTATACAGGAGCGCTTGCCAAATATCTGCA 112
Qy 1382 ATATGGGAGTGGATGTGAAGCAAACTGTTTCAGGCATTGAACAAGCTTGTTCGATGTAAT 1441
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
111 ATATGGGAGTGGATGTGAAGCAAACTGTTTCAGGCATTGAACAAGCTTGTTCGATGTAAT 52
Qy 1442 GATTTCGAAAACAATGTGATATTCGACTTTAAAAATATCAAAAGTTAATTC 1492
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
51 GATTTCGAAAACAATGTGATATTCGACTTTAAAAATATCAAAAGTTAATTC 1

Search completed: May 29, 2003, 17:06:25
Job time : 2094 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 17:06:35 ; Search time 306 Seconds
(without alignments)
3429.514 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFQILVFLHALLIF.....CNMGVDVQKTVSAIEQACSM 466

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US0934066/runat_19052003_163659_24347/app_query.fasta_1.647
-DB=N_Geneseq_101002 -QWMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0934066_@CGN_1_1_396_@runat_19052003_163659_24347 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2454	100.0	1401	21	AAC50984
2	2446	99.7	1541	21	Arabidopsis thalia
3	1268	51.7	1742	14	Asparaginylendopep
4	1265	51.5	1749	14	Asparaginylendopep
5	1251	51.0	1323	14	Asparaginylendopep
6	1246	50.8	1485	14	Arabidopsis thalia
7	1245	50.7	1323	21	Asparaginylendopep
8	1220	49.7	1830	14	Asparaginylendopep
9	1215	49.5	1323	14	Asparaginylendopep
10	1080	44.0	1323	14	Asparaginylendopep
11	1080	44.0	1323	14	Asparaginylendopep
12	1080	44.0	1910	14	Asparaginylendopep
13	1019	41.5	1640	14	Asparaginylendopep
14	990	40.3	1152	14	Asparaginylendopep
15	891	36.3	894	14	AAO50565
16	873.5	35.6	813	14	AAO50566
17	828	33.7	1936	19	AAV44886
18	828	33.7	2030	22	AAC68662
19	805.5	32.8	753	14	AAQ50569
20	771	31.4	1128	19	AAV29038
21	771	31.4	1749	19	AAV29037
22	749	30.5	1453	14	AAO50573
23	744	30.3	1855	18	AAV66381
24	744	30.3	1855	20	AAV69241
25	744	30.3	1855	21	AAZ44095
26	685	27.9	1336	24	ABK48384
27	629.5	25.7	2373	23	AA565199
28	613	25.0	1250	14	AAO50571
29	326.5	13.3	981	23	ABL07607
30	300.5	12.2	3172	23	ABL07606
31	269	11.0	1490	21	AAC36486
32	254.5	10.4	1491	21	AAC49045
33	227.5	9.3	423	16	AACT35112
34	216.5	8.8	640	23	AA565197
35	195	7.9	2135	22	AAC83228
36	178	7.3	220	18	AAV66394
37	178	7.3	220	20	AAV69258
38	178	7.3	220	21	AAZ44108
39	168	6.8	219	18	AAV66390
40	168	6.8	219	20	AAV69254
41	168	6.8	219	21	AAZ44104
42	162.5	6.6	461	21	AAQ00148
43	148	6.0	329	14	AAQ50578
44	146	5.9	15061	22	AAI04226
45	146	5.9	15061	23	ABK42104

ALIGNMENTS

RESULT 1
AAC50984
ID AAC50984 standard; DNA; 1401 BP.
XX
AC AAC50984;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 66846.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX


```

PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 2,27e-231 Length: 1401
Score: 2454.00 Matches: 466
Percent Similarity: 100.00% Conservations: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-934-066-2 (1-466) x AAC50984 (1-1401)

Qy 1 MetSerProLeuGlyHisPheGlnLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 1 ATGCTAGTCTCTTGGTCACTTTTCAGATTCTTTTCTTCTCATGCTTTTCTTATCTTC 60

Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAspValGluSerSerAspLys 40
Db 61 TCAGCTGAGTCCGCCAAACCCCAATTCCTGAACGATAATGATGTGAATCTAGCGCAAG 120

Qy 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyTyTyAsnTyr 60
Db 121 AGTCAAAAGGCACACACATGGGCTGTTTGTAGTCTGATCAAAATGATATATAACTAC 180

Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyTyGlnIleLeuArgLysGlyGlyLeuLysAsp 80
Db 181 AGGCATCAGGCTGACATATGCCACGCTATCAGATACCTCCGAAAGCGGTTTAAAGAT 240

Qy 81 GluAsnIlelleValPheMetTyTyAspIleAlaPheSerSerGluAsnProArgPro 100
Db 241 GAAACATCATATGTTGTTTATGTATGATGATATCCGTTTCTCCGAGAAATCTTAGCCCT 300

Qy 101 GlyValIlelleAsnLysProAspGlyGluAspValTyTyLysGlyValProLysAspTyr 120
Db 301 GGAGTTATCATTAATAAACCATGGAGAGATGTTTATAAGGAGTTCCTTAAGGACTAC 360

```

RESULT 2

```

Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyTyAsnValLeuLeuGlyAsnGluSerGly 140
Db ACTAAAGAAAGCTGTTAATGTTCAAAACTTCTACAATGTGTACTTGGAAATGAAAGTGC 420

Qy 141 ValThrClyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db GTACAGGAGGAAATGGCAAGTGTGAAAAGTGGTCTTAATGATAATATCTTCATCTAT 480

Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db TATGCTGACCATGGAGCTCTGCTTAATAGCATGCCACTGGTATGAAGTATATGCA 540

Qy 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyTyAsnLysMetVal 200
Db AAAGATTCAATGAAGTCTTTGGAGAGATGCATAAGAGAAAAAATAACAAGATGCTG 600

Qy 201 IleTyTyValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db ATCTATGTTGAGCATGTGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAAATCTC 660

Qy 221 AsnIleTyTyAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyTyCysPro 240
Db AACATATACGACGACTGCTGCTAATTTCTAAAGAGACAGCTGGGAGTTCCTACTGCT 720

Qy 241 GluSerTyTyProProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db GAGTCATATCTCTCTCTCTCTGAGATTGGAACCTTCTCTCGGCGATACATTTAGCATC 780

Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
Db TCTTGGCTTGAGGACAGTGCCTTCATGACATGACAAAGAGACTTTTGGAGCAACAATAC 840

Qy 281 HisValValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db CACGTTGTAAGACAGACAGTAGGATCTGATGTACCAGACTTCTCATGTATGCGGTTTC 900

Qy 301 GlyThrGluLysMetLeuLysAspTyTyLeuSerSerTyTyIleGlyArgAsnProGluAsn 320
Db GGAACAGAGAAAGTGTAAAGATTATCTTTCTCTTACATTGGAAGAAATCTCTGAAAC 960

Qy 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db GATAACTTCACTTTTCACGGAATCTTTTCTTCAACAATCTCTAATCTGCTTGGTCAAT 1020

Qy 341 ProArgAspIleProLeuLeuTyTyLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db CCGCGCGATATTCCTCTGCTATACCTCCAGAGAAAGATTCAAAAAGCTCCCAATCGGATCA 1080

Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db CTTGAAAGCAAAAGAGCTCAGAGAAATGCTTGACGAAAGAAATCATAGGAAACAATC 1140

Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db GATCAGAGCATTCACAGACATTCCTCGGCTTTCAGTTAAACAAACCAATGTCTTAAATCTC 1200

Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db TTAACCTTCCACAGACAAACAGACAGACGCTTCTTGTAGACCATTTGGGATGCTTCAAGACT 1260

Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyTyGlyLeuLysTyThr 440
Db CTAGTTAATAGTCTCAAGAAATCACTGCGGTGCAACGCTGCATTTACGATTTGAAGTATACA 1320

Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db GGAGCGCTTCCCAATATCTGCAATATGGGAGTGGATGCTGAAGCAAACTGTTTACGCCATT 1380

Qy 461 GluGlnAlaCysSerMet 466
Db 1381 GAACAAGCTTGTTCGATG 1398

```

AAC35910
ID AAC35910 standard; DNA: 1541 BP.
XX AC AAC35910;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 11859.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;
XX KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 28-JUL-1999; 99US-0145319.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspSerTrpAspCysPheLysThr 420
 Db 1259 TTAACCTCCACAAGAACACAGCAGCCTCTGTGTAGACGATTGGGATTGCTTCAAGACT 1318
 Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
 Db 1319 CTAGTTAATAGCTTCAAGAACTACCTGCGTGCACGGTGCATTACGGATTGAAGTATACA 1378
 Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
 Db 1379 GGAGCGCTTGCCTCAATATCTGCAATATGGAGTGGATGTGAAGCAAACTGTTTCAGCCATT 1438
 Qy 461 GluGlnAlaCysSerMet 466
 Db 1439 GAACAGCTTGTCTCGATG 1456

RESULT 3
 AAQ50570
 ID: AAQ50570 standard; cDNA to mRNA; 1742 BP.

AC AAQ50570;

XX 24-MAY-1994 (first entry)

XX Asparaginyl endopeptidase clone 101.

XX Asparaginyl endopeptidase; Canavalia ensiformis; seed;

KW L-asparagine; primer; PCR; protein fragmentation;

KW peptide synthesis; ss.

OS Canavalia ensiformis.

XX Key Location/Qualifiers

FT CDS 207..1529

FT /*tag a

XX JP05276960-A.

XX 26-OCT-1993. 92JP-0231602.

XX 07-AUG-1992; 92JP-0056023.

XX (SYUZ/) SYUZO T.

XX WPI: 1993-373587/47.

XX P-PSDB: AAR43035.

XX New gene for encoding asparaginyl endo-peptidase - comprises 8

XX specified DNA sequences

XX Disclosure: Page 17-19; 35pp; Japanese.

XX A gene encoding asparaginyl endopeptidase is claimed.

CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease

CC derived from a seed of Canavalia ensiformis which selectively

CC hydrolyses C-terminus amide bond of L-asparagine residue

CC (see AAR43033 and AAR43041).

CC The enzyme is useful for protein fragmentation and enzymatic

CC peptide synthesis.

CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used

CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,

CC by PCR.

XX Sequence 1742 BP; 507 A; 338 C; 395 G; 502 T; 0 other;

SQ Alignment Scores:

Pred. No.: 1,068-114 Length: 1742

Score: 1268.00 Matches: 243

Percent Similarity: 65.97% Conservative: 71

Best Local Similarity: 51.05% Mismatches: 140

Query Match: 51.67% Indels: 22

DB: 14 Caps: 4
 US-09-934-066-2 (1-466) x AAQ50570 (1-1742)
 Qy 2 SerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuPhe--- 20
 Db 75 TCAATTACCATGAACCCCTTTTCGATCCTCTTCGATCCGCCACCTCTCTGCTCTTCGTC 134
 Qy 21 -----SerAlaGluSerArgLysThrGlnLeu 29
 Db 135 TCCGGTGACCATGACGATGTTCTCCGATTGCCCTCCGAAGCTTCCAGATCTTCCAAAGCA 194
 Qy 30 LeuAsnAspAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTrpAlaVal 49
 Db 195 CCCGAGGAAACGAC-----GAGGGGACAGGTGGCCGTT 230
 Qy 50 LeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAla 69
 Db 231 TTAATTGCTGTTCCCAATGGTTACTGGAATTACAGGCATCAGTCTGATGTATGCCATGG 290
 Qy 70 TyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyrAsp 89
 Db 291 TATCAAGTGTGAGGAAAGGTGCTGAAAGAAGAAATATTGTTGTTTATGTATGAT 350
 Qy 90 AspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAspGly 109
 Db 351 GACATTGCTTTCCATAAAGAAACCCACGCGCTGGAATCATCTTAACAGTCCACATGGA 410
 Qy 110 GluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnValGlnAsn 129
 Db 411 GATGATGTTTACAAGGAGTACCTTAAGGATTACGTTGGTAAAGATGTTACTGTTTANCAAC 470
 Qy 130 PheTyrAsnValLeuLeuGlyAsnGluSerGlyValIleThrGlyLysAsnGlyLysValVal 149
 Db 471 TTTTGTAGCTGCTACTTGGAAATAAGTCAAGCTATTACCGCGGTAGTGGGAGGTGTC 530
 Qy 150 LysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeu 169
 Db 531 AATAGTGTCCCAATGATCATATATTTATTTACTCTCATCATGGGCGTCCGGAGTG 590
 Qy 170 IleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuGlyLys 189
 Db 591 CTGGGATGCTCTACTAGTCCGTTCTGTATGCTACTGATCTGATGAAGCTTTGAAGAAG 650
 Qy 190 MethisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGly 209
 Db 651 AAGCATGCTTCTGGAAACATATAAAGCCTTGTATTTTATCTAGAGGCATGTCAATCTGG 710
 Qy 210 SerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsn 229
 Db 711 AGTATCTTTGAAGCTCTTCTCTGGAAGGTCTAAATATCTATGCAACAACAGCTCAAAAT 770
 Qy 230 SerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProProSerGlu 249
 Db 771 GCAAAAGAAACGACTTGGGAAACATATTTCCTCTGGGAGTCTCTAGTCTCTCCCCAGAA 830
 Qy 250 IleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHis 269
 Db 831 TATGAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 890
 Qy 270 AspMetSerLysGluThrLeuGluGlnIleTyrHisValValLys---ArgArgValGly 288
 Db 891 AATTATCAACAGAAACCTCTACATCAACATACGAATCGGTCAAGCAAGGACTATCAAT 950
 Qy 289 SerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAsp 308
 Db 951 CGAAATTCATTTTATGTTCTCAGTGATGTCATGTTGGTACATAGATAGGCTTTAGCAAAAAT 1010
 Qy 309 TyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSer 328
 Db 1011 AATCTCTTCTATATTTGGGTGCAAAATCTCTGCAAAATCATAAATTTTACTTTTCTGATAAA 1070
 Qy 329 PheSerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyr 348

Db 964 AGGACCATAGTGGAGTTTCATATGCTCTCAGCTGATGCGAGTATGCTATAGG 1023
 Qy 305 MetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThr 324
 Db 1024 CACGGATGATGTTCTCTCTCTATATTTGGGTACAATCTGCTAATGATAATTTTACC 1083
 Qy 325 PheThrGlu-----SerPheSerSerProIleSerAsnSerGlyLeuValAsnProArg 342
 Db 1084 TTGTGCGATGAACACTCTCTGAGGTCACCT-----TCMAAAGCAGTCAACCAACG 1134
 Qy 343 AspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGlu 362
 Db 1135 GATGCTGATCTCATCTATTTCTGGGAGAGTTCGCAAGCTCTGAGGTTCTCCGAG 1194
 Qy 363 SerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGln 382
 Db 1195 AAAATGCTGCTCAGAAACAGTTCTGGAAGTAATGTCTCAGAGGATGCATATAGACAAC 1254
 Qy 383 SerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuThr 402
 Db 1255 GGTGTGGAACCTATTGTTGGAGCTTTTATTGTCATTTGAAAGGTCCAAAGTACTGGAT 1314
 Qy 403 SerThrArgThrThrClyClnProLeuValAspAspTrpAspCysPheLysThrLeuVal 422
 Db 1315 GCTGTTAGACCGCTGGAATGCGCACTTGCTGACTGGGACTCCCTGAAACCACTGGTA 1374
 Qy 423 AsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAla 442
 Db 1375 AGGACATTTGACACATATTTGTTGGATCCTTGCTCAGTATGCGATGGAACATATGAGGTCC 1434
 Qy 443 LeuAlaAsnIleCysAsnMetGlyValAspValLysClnThrValSerAlaIleGluGln 462
 Db 1435 TTGCAACATCTGCAACGACGGAATTAAGAATGACCAAAATGGCCGATGCTCAGCACAA 1494
 Qy 463 AlaCys 464
 Db 1495 GCTTGT 1500

RESULT 5

ID AAO50562 standard; cDNA to mRNA; 1323 BP.

AC AAO50562;

DT 24-MAY-1994 (first entry)

DE Asparaginylendopeptidase sequence (4).

KW Asparaginylendopeptidase; Canavalia ensiformis; seed;

KW L-asparagine; primer; PCR; protein fragmentation;

KW peptide synthesis; ss.

OS Canavalia ensiformis.

PN JP05276960-A.

PD 26-OCT-1993.

PF 07-AUG-1992; 92JP-0231602.

PR 07-FEB-1992; 92JP-0056023.

PA (SYUZ/) SYUZO T.

DR WPI; 1993-373587/47.

PT New gene for encoding asparaginyl endo-peptidase - comprises 8

PT specified DNA sequences

PS Claim 1; Page 13-14; 35pp; Japanese.

CC A gene encoding asparaginylendopeptidase is claimed.

CC 8 DNA sequences are given (AAO50559-66). The enzyme is a protease

CC derived from a seed of Canavalia ensiformis which selectively
 CC hydrolyses C-terminus amide bond of L-asparagine residue
 CC (see AAR43033 and AAR43041).
 CC The enzyme is useful for protein fragmentation and enzymatic
 CC peptide synthesis.
 CC The primers given in AAO50567-68, AAO50576-77 and AAO50583-90 were used
 CC in the isolation of the fragments given in AAO50569-75 and AAO50578-79,
 CC by PCR.
 XX Sequence 1323 BP; 375 A; 255 C; 328 G; 365 T; 0 other;
 SQ

Alignment Scores:

Pred. No.: 3,43e-113 Length: 1323
 Score: 1251.00 Matches: 233
 Percent Similarity: 70.05% Conservative: 64
 Best Local Similarity: 54.95% Mismatches: 121
 Query Match: 50.98% Indels: 6
 DB: 14 Gaps: 3

US-09-934-066-2 (1-456) x AAO50562 (1-1323)

Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
 Db 7 GGTACCAAGGTGGGCCATCTTCGCGCGTTCCTCAATGGCTACTGGAATTACAGGCATCAG 66
 Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIle 83
 Db 67 GCTGATATTTGTCATGCGTATCAACACACTCAGGAAAGTGGCCTGAAAGAGAAATATT 126
 Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
 Db 127 ATTTCTTTTATGATGATGACATTTGCTTCAATGAGGAGAACCCAGACCTGGGTCATC 186
 Qy 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
 Db 187 ATTAACAACACAGATGGGACTGATGTTTATGAAGGAGTTTCCAAAGGATTACACTGCCAA 246
 Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
 Db 247 GATCTTACTGCTGNAACCTTCTACGCTGTTTACTTGGAAATAAGTCGCCACTTAGCGGT 306
 Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
 Db 307 GGCATGGGAGGTTTGAAACAGTGGCCCGCATGCTGATATTTGATTTCTATAGTACT 366
 Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
 Db 367 CATGGAGTCCAGAGTTCTCGGGATGCTTCCGCGTCTTACTTATATGCTATCTGATCTG 426
 Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
 Db 427 AATGATCTTGAAGAAAAACATGCTCCGGAACATATATAAAGCCTAGTATTTATCTT 486
 Qy 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
 Db 487 GAGCATGCTGAATCTGGAGTATCTTTTGAAGGCTTCTTCTCTGAAGATGCTAATATTTAT 546
 Qy 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
 Db 547 GCAACAACAGCTTCGAATGCAAGAAAGACAGCTGGGGAACATATTTGCCAGGGAGTAT 606
 Qy 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
 Db 607 CCCAGTCTCTCCCGAGANTATTCACTTCTTGGGAGACCTGTACAGTGTCTGGGATG 666
 Qy 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlyThrHisVal 283
 Db 667 GAACAGATGCACAGACAATTTGCCAACTTGAACCACTTTGAACCAACAATATAAATTTGGT 726
 Qy 284 LysArgArg--ValGlySerAspValProGluThrSerHisValCysArgPheGlyThr 302
 Db 727 AAGAGAGGAGCAATTAGTGGAGTTTCATCTACTATGCTCTCAGCTGATGCAGTATGCTGAT 786

```

Qy 303 GluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsn 322
Db 787 ATAGGGCTCAGCGATGTTCTCTCTATTTGGGTACAATCTGCTAATGATAAT 846
Qy 323 PheThrPheThrGlu-----SerPheSerSerProfileSerAsnSerGlyLeuValAsn 340
Db 847 TTTACCTTTGCGATGAAACTCTTGAGGTCACCT-----TCAAAAGCAGTCAAC 897
Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db 898 CAACGGGATCCTCATCTCATCCATTTCTGGGAGAAGTTGCGCAAGCTCCTGAGGGTTCT 957
Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 958 CCCGAGAAAAATCTGCTCAGAACCAAGTTGTGGAAGTAATGCTCACAGGATGCATATA 1017
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1018 CACAACGGTGTGAACCTTATTGGAGCTTTATTGGCATTGAAAGGGTCCAAAAGTA 1077
Qy 401 LeuThrSerThrArgThrThrGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db 1078 CTGATGCTCTTAGACGGCTGGAATGGCACCTGTTGATCAGCTGGGACTGCTCGTGAACCC 1137
Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db 1138 ATGCTAAGGACATTTGAGACATATTGTCGATATTTGTCAGTATGGGATGAACATATG 1197
Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db 1198 AGTCTCTTCCAACATCTCCAACGGCAGGAATTAGAATGACCAATGGCCGATGCCTCA 1257
Qy 461 GluGlnAlaCys 464
Db 1258 GCACAAGCTTGT 1269

RESULT 6
AAC42934
ID AAC42934 standard; DNA; 1485 BP.
XX
AC AAC42934;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37409.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.

```



```

Db 674 TATTGCCCGAGGAGGATCCCTCTCCCCAGAAATATTCACATGCTTGGTGACGAG 733
Qy PheSerIleSerTrpLeuGluAspSerLeuHisAspMetSerLysGluThrLeuGlu 277
   :::::::::::::::::::::
Db 734 TACAGTGTGCTGGATGAGACACATGACAGGAAGAAATTTGGACAGAACTTTGCAC 793
Qy GlnGlnTrpHisValValLysArgArg---valGlySerAspValProGluThrSerHis 296
   :::::::::::::::::::::
Db 794 CACACATATGAATTTGGTTAAAGAGAGAGACTATTAAACGGAAGTATATACCATAGCTCTCAC 853
Qy ValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArg 316
   :::::::::::::::::::::
Db 854 GTGATCGCATGATGGTATTAAGTCTACAGGATGATGTTCTCTCTCTATATTTGGGTACA 913
Qy AsnProGluAsnAspAsnPheThrPheThrGlu-----SerPheSerSerProIleSer 334
   :::::::::::::::::::::
Db 914 AATCCTGCTAATGATAATTTTACCTATGTGGTGAAGTCTCCTTGAGGTCACT----- 967
Qy AsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGln 354
   :::::::::::::::::::::
Db 968 ---TCAAAAGCAATCAGCCAACTGTTGCTGATCTCATCCATTTTGGGAGAGTTCCGC 1024
Qy LysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLys 374
   :::::::::::::::::::::
Db 1025 AAGCTCTCTGAGGTTCTACAGGAAGAGTGTGCTCTCAGAAACAATTTCTGGAAGTAAG 1084
Qy AsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGln 394
   :::::::::::::::::::::
Db 1085 TCTCACAGATGATATAGACAACACTGTCAAAATTTATGGGAGTCTTTTATTTGGCAAT 1144
Qy ThrAsnValLeuAsnLeuThrSerThrArgThrThrGlyGlnProLeuValAspAsp 414
   :::::::::::::::::::::
Db 1145 GAAAGGGTCCAGAGTACTCAATGCTGTAGACCGCTGGAATGGCACTTGTGTATGAC 1204
Qy TrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHis 434
   :::::::::::::::::::::
Db 1205 TGGGACTGCTGAAAATATGTTAAGGACTTTTGAGACATATTTGGTGGCTTGTCTCAG 1264
Qy TyrGlyLeuLysThrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLys 454
   :::::::::::::::::::::
Db 1265 TATGGGATGAATATATATAGGTCCTTTTCAAAACATCTGCAATCAAGAAATTAAGATGAC 1324
Qy GlnThrValSerAlaIleGluGlnAlaCys 464
   :::::::::::::::::::::
Db 1325 CAAATGGCTGATGCCCTCAGCACAAAGCTTGT 1354

RESULT 9
ID AAQ50561 standard; cDNA to mRNA; 1323 BP.
XX
AC AAQ50561;
XX
DT 24-MAY-1994 (first entry)
XX
DE Asparaginylendopeptidase: sequence (3).
XX
KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
KW L-asparagine; primer; PCR; protein fragmentation;
KW peptide synthesis; ss.
XX
OS Canavalia ensiformis.
XX
PN JP05276960-A.
XX
PD 26-OCT-1993.
XX
PF 07-AUG-1992; 92JP-0231602.
XX
PR 07-FEB-1992; 92JP-0056023.
XX
PA (SYUZ/) SYUZO T.
XX
WP 1993-373587/47.

```

```

XX New gene for encoding asparaginyl endo-peptidase - comprises 8
PT specified DNA sequences
XX
PS Claim 1; Page 13; 35pp; Japanese.
XX
CC A gene encoding asparaginylendopeptidase is claimed.
CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
CC derived from a seed of Canavalia ensiformis which selectively
CC hydrolyses C-terminus amide bond of L-asparagine residue
CC (see AAR43033 and AAR43041).
CC The enzyme is useful for protein fragmentation and enzymatic
CC peptide synthesis.
CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
CC by PCR.
XX
SQ Sequence 1323 BP; 390 A; 244 C; 312 G; 377 T; 0 other;

Alignment Scores:
Pred. No.: 1-18e-109 Length: 1323
Score: 1215.00 Matches: 226
Percent Similarity: 68.63% Conservative: 65
Best Local Similarity: 53.30% Mismatches: 127
Query Match: 49.51% Indels: 6
DB: 14 Gaps: 3

US-09-934-066-2 (1-466) x AAQ50561 (1-1323)
Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
Db 7 GGTACCAGATGGGCCATCTTAATCGCGCTCCCAATGGTACTGGAAATTACAGGCATCAG 66
Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
Db 67 GCTGATATTGTGTCGCTGATCAATGACTGAGGAAGGTGGCCCTGAAAGAAATAATT 126
Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
Db 127 ATGTGTTTATGATCATGACATGCTTCAATGAGAACCAACCAACCTGGAGTCATC 186
Qy 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
Db 187 ATTAACAACCCAGATGGGATGATGTTTATGAGGAGTTCCTCAAGAGGATTACATCGCGC 246
Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
Db 247 GATGTTACTCTCATAACTTCTATGCTGTATATCTTGGAAATAAATCAGCAGCTTACAGT 306
Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
Db 307 GGCAGTGGGAAGGTTGTAACAGTGGCCCTGATGATCGTATATTCATATTCATATAGTGAT 366
Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
Db 367 CATGGAGGTCCAGAGTCTTCTGGGAGTGGCTGGTGGCTTCTATATGATCATGATCTG 426
Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 427 ATTGAAGCTTTGAAGAAAACATGCTTCTGGAGACGTATATAAAGCCCTTGTATTATCTT 486
Qy 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
Db 487 GAGGCATGTGAATCTGGAGATCTTTTGAAGGCTTCTTCTCTCAAGATATCAATATTTAT 546
Qy 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
Db 547 GCAACAACAGCTTCCCAATGCAGAAAGAACAGCTGGGGAACATATTTCCCGCAGGAGAT 606
Qy 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
Db 607 CCCAGTCTCCCGCCAGCAATATTCACCTTCTGCTGGTGACGAGTACAGTGTGCTGGATG 666

```

```

QY 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal 283
DB 667 GAAGACAGTGCAGCAAGATTTGGACACAACTTTGACCAACATATGAATGGTT 726
QY 284 LysArgArg---ValGlySerAspValProGluThrSerHisValCysArgPheGlyThr 302
DB 727 AAAGACAGAGCATATTATACCGAAGATATAPCCATGCTCAGCTGATGCGAGTATGGTAT 786
QY 303 GluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsn 322
DB 787 ATAAGTCTCAGCGATGATGTTCTCTCTATATTTGGGTACAAATCTGCTGATAAT 846
QY 323 PheThrPheThrGlu-----SerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
DB 847 TTACCTATGTGGATGAGAACTCCTTGAGTGCACCT-----TCAAAAGCAATCAGC 897
QY 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
DB 898 CAAGGTGTTGCTGATCTCATCCATTTTGGGAGAAGTTCGCAAAAGCTCCTGAGGTTCT 957
QY 361 LeuGluSerLysGluAlaGlnLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
DB 958 ACCAGGAAGATGCTGCTCAGAACAAATTTCTGGAAGTATGCTCAGCAATGTCATATA 1017
QY 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
DB 1018 GACACAGTGTGAATAATTATTTGGGAGTCTTTTATTTGGCATTTGAAAGGCTCCAGAGTA 1077
QY 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
DB 1078 CTCATGCTGTGTAGACCGCTGGAATGGCACTTTGTCATGCTGCTGCTGCTGCTGCTG 1137
QY 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
DB 1138 ATGGTAAGGACTTTTGAGACATATTTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
QY 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
DB 1198 AGGTCTTTTGCACAACTCTGCAATGCAAGTAATAAGAAATGACCAATGCTGCTGCTGCT 1257
QY 461 GluGlnAlaCys 464
DB 1258 GCACAGCTTGT 1269

RESULT 10
AAQ50559
ID AAQ50559 standard; cDNA to mRNA; 1323 BP.
XX
AC AAQ50559;
DE
DT 24-MAY-1994 (first entry)
XX
DE Asparaginylendopeptidase sequence (1).
XX
KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
KW L-asparagine; primer; PCR; protein fragmentation;
KW peptide synthesis; ss.
XX
OS Canavalia ensiformis.
XX
JP05276960-A.
XX
PN
XX
PD 26-OCT-1993.
XX
PF 07-AUG-1992; 92JP-0231602.
XX
PR 07-FEB-1992; 92JP-0056023.
XX
PA (SYU2/) SYU20 T.
XX
DR WPI; 1993-373587/47.
PT New gene for encoding asparaginyl endo-peptidase - comprises 8

```

```

PT specified DNA sequences
XX Claim 1; Page 12; 35pp; Japanese.
CC
XX A gene encoding asparaginylendopeptidase is claimed.
CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
CC derived from a seed of Canavalia ensiformis which selectively
CC hydrolyses C-terminus amide bond of L-asparagine residue
CC (see AAR43033 and AAR43041).
CC The enzyme is useful for protein fragmentation and enzymatic
CC peptide synthesis.
CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
CC by PCR.
XX
SQ Sequence 1323 BP; 385 A; 246 C; 346 G; 346 T; 0 other;

Alignment Scores:
Pred No.: 2,18e-96 Length: 1323
Score: 1080.00 Matches: 207
Percent Similarity: 65.278 Conservativity: 73
Best Local Similarity: 48.258 Mismatches: 129
Query Match: 44.018 Indels: 20
DB: 14 Gaps: 4

US-09-934-066-2 (1-466) x AAQ50559 (1-1323)
QY 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
DB 7 GGAACCCGGTGGCGGTTCTCGTGGCTGTTCAACGGTTATGGAAATTTATAGGCATCAA 66
QY 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
DB 67 CCGGATGTGTCATGCTTACCAAGTTCATAAAAGTGGAGTCAAGAGCAAAACATT 126
QY 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
DB 127 GTGCTGTTTATGATGATATAGCTTATACGCCATGAATCCAGACCCCGAGTCATC 186
QY 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
DB 187 ATCAACCATCTCAGGGCCAGACGCTGATGCTGTGTACCTAACGNTTACACCGGTAG 246
QY 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
DB 247 GACGTAAACCTGAGAACCTATATGCTGCTCTTGGGACAAAGATAAAGTTAAAGT 306
QY 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
DB 307 GGAAGTGGCAAGGTGATCAACCAATCCCGAGGATAGGATATTTATATCTACTCCGAT 366
QY 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
DB 367 CATGGAGTCCCGAGTTCCTGGGATGCCAACGCCACCATTCGTTTATGCCATGATTTT 426
QY 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
DB 427 ATTGATGTTTGAAGAAGAACATGCAAGTGGAGCTACAAGCAGATGCTTATACATA 486
QY 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
DB 487 GAAGCTGTGAGAGTGGAGCATATTTGAGGTATATGCCCCAAGGATCTGAATTTAT 546
QY 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
DB 547 GTGACAACTCGCTCAATGCACAGAACAGATTTTGGAACTTATTTGCTCCGGATGAAT 606
QY 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
DB 607 CCTCTCCACACAGAGAGTACGTAACCTGCTGGGGATTTATACAGCGTTTCTTTGGATG 666
QY 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal 283
DB

```

667	GAGATAGT	CAGACT	CACAATCT	TAAGAAGGAACGGTACAACAGCAANTACCAGT	TCGGTA	726
284	LysArgArgVal	GlySerAspVal	ProGluThr	-----SerHisValCysArgPhe	Gly	301
727	AGGAACGGACTT	CAAAATCT	TAACAGCTAT	AGGTTTGGTTCTCATGTGATGCAATACGGT		786
302	ThrGluLysMet	LeuLysAspTyr	LeuSerSerTyr	IleGlyArgAsnProGluAsnAsp		321
787	GACACTAACAT	TACTGCTGAAAGCTT	TACTTGTC	TACCATGGTTTGTATCTCCTCCACCGCTG		846
322	AsnPheThr	PheThrGluSerPheSer	ProIleSerAsnSer	GlyLeu-----		338
847	AACTTT-----					
339	-----Val	AsnProArgAsp	IleProLeuLeuTyr	IleGlnArgLysIleGlnLys		355
883	ATGGAAGTTGTTT	TAACAGAGAGATG	CAGAGCTTTT	TGTTTCATGTGCGCAAAATGTATCAGAGA		942
356	AlaProMet	GlySerLeuGluSerLys	GluAlaGlnLysLysLeuLeuAsp	GluLysAsn		375
943	TCAAACCAT	CAACCGGAAAGACAG	ACTCACA	CTCTGGAACAGATTACAGACAGACAGTGAAG		1002
376	HisArgLysGln	IleAspGlnSer	IleThrAsp	IleLeuArgLeuSerValLysGlnThr		395
1003	CATAGGAATCAT	TTGGATGCGCAGT	GCGAATTGAT	TGCGAGCTTTTGTGTTATGGACCAGGA		1062
396	AsnValLeu	AsnLeuLeuThrSer	ThrArgThrThr	GlyGlnProLeuValAspAspTrp		415
1063	AAAAGTTCTCGGTT	CTTACATTC	CGTGAGGCGCTC	TGGTCTGCCCCCTAGTTGATGATTGG		1122
416	AspCysPhe	LysThrLeuValAsnSer	PheLysAsnHisCys	GlyAlaThrValHisTyr		435
1123	ACATGCTTGA	AATCTATG	GGTTAGAGTGTTCG	AAACTCACTGTGGTCACTGACTCAGTAT		1182
436	GlyLeuLysTyr	ThrGlyAlaLeuAlaAsn	IleCysAsnMet	GlyValAspValLysGln		455
1183	GGCATGAACACAT	CGGGCATTCG	GCACAGTTTGCAAC	AGCGCGTT-----		1230
456	ThrValSer	AlaIleGluGlnAlaCys	464			
1231	TC	TAAAGCCCTCCATG	GAGGAGGCTGT	1257		
RESULT 11						
AAQ50563						
ID	AAQ50563 standard; cDNA to mRNA; 1323 BP.					
XX	AAQ50563:					
XX	24-MAY-1994 (first entry)					
XX	Asparaginylendopeptidase sequence (5).					
DE	Asparaginylendopeptidase; Canavalia ensiformis; seed;					
KW	L-asparagine; primer; PCR; protein fragmentation;					
KW	peptide synthesis; ss.					
XX	Canavalia ensiformis.					
OS	JP05276960-A.					
XX	26-OCT-1993.					
XX	07-AUG-1992; 92JP-0231602.					
XX	07-FEB-1992; 92JP-0056023.					
XX	(SYUZ/) SYUZO T.					
XX	WPI; 1993-373587/47.					
XX	New gene for encoding asparaginyl endo-peptidase - comprises 8					
PT	specified DNA sequences					
XX						


```

Db 895 GAGATAGTGAGACTCAATCTAAAAAGGGAACGGTACAACAGCAATACCAGTCGGTA 954
Qy 284 LysArgArgValGlySerAspValProGluThr-----SerHisValCysArgPheGly 301
Db 955 AGNAGAGGACTCAAAATCTAACAGCTATAGGTTGGTTCTCATGTGATGCAATACGGT 1014
Qy 302 ThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAsp 321
Db 1015 GACACTAACATTACTGCTGAAAAGCTTTACTGTGACCATTGTTGATCTCTGCCACCCTG 1074
Qy 322 AsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeu----- 338
Db 1075 AACTTT-----CCTCCACACACAGCGCAACCTAGAGCTAA 1110
Qy 339 -----ValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLys 355
Db 1111 ATGGAAGTTGTTAACAGAGAGATGACAGACTTTTGTTCATGTGGCAATGTATCAGAGA 1170
Qy 356 AlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsn 375
Db 1171 TCAAAACCATCAACCGGAAAAGAGACTCACATCTGGAACAGATTACAGAGACAGTGAAG 1230
Qy 376 HisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThr 395
Db 1231 CATAGGAATCATTTGGATGACAGCTGTGGAATGATTGAGCTTTTGTGTATGGACACGGA 1290
Qy 396 AsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTyr 415
Db 1291 AAAAGTTCTTCGGTTCTACATCCGTGAGGCTCGTGGTCTGCCCTAGTTGATGATGG 1350
Qy 416 AspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyr 435
Db 1351 ACATGCTTGAATCTATGTTAGAGTGTTCGAAACTCACTGTGGTCACTGACTCACTAT 1410
Qy 436 GlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAlaSpValLysGln 455
Db 1411 GGATGAACACATGCGGGCATTCGGCAACGTTTGGCAACAGCGCGT----- 1458
Qy 456 ThrValSerAlaIleGluGlnAlaCys 464
Db 1459 TCTAAGGCTCCATGGAGAGGCTTGT 1485

RESULT 13
ID AAQ50575 standard; CDNA to mRNA; 1640 BP.
AC AAQ50575;
XX
XX
XX 24-MAY-1994 (first entry)
XX
XX Asparaginylendopeptidase clone ASN-1.
XX
XX Asparaginylendopeptidase; Canavalia ensiformis; seed;
KW L-asparagine; primer; PCR; protein fragmentation;
KW peptide synthesis; ss.
XX
XX Canavalia ensiformis.
XX
XX JP05276960-A.
XX
XX 26-OCT-1993.
XX
XX 07-AUG-1992; 92JP-0231602.
XX
XX 07-FEB-1992; 92JP-0056023.
XX
XX (SYUZ/) SYUZO T.
XX
XX WPI; 1993-373587/47.
XX
XX New gene for encoding asparaginyl endo-peptidase - comprises 8
PT specified DNA sequences

```

```

xx
PS Disclosure; Page 26; 35pp; Japanese.
xx
CC A gene encoding asparaginylendopeptidase is claimed.
CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
CC derived from a seed of Canavalia ensiformis which selectively
CC hydrolyses C-terminus amide bond of L-asparagine residue
CC (see AAR43033 and AAR43041).
CC The enzyme is useful for protein fragmentation and enzymatic
CC peptide synthesis.
CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
CC by PCR.
xx
SQ Sequence 1640 BP; 533 A; 288 C; 383 G; 436 T; 0 other;

Alignment Scores:
Pred. No.: 2.88e-90 Length: 1640
Score: 1019.00 Matches: 195
Percent Similarity: 64.42% Conservative: 73
Best Local Similarity: 46.88% Mismatches: 128
Query Match: 41.52% Indels: 20
DB: 14 Gaps: 4

US-09-934-066-2 (1-466) x AAQ50575 (1-1640)
Qy 57 TyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGly 76
Db 4 TATGGTAATTATCGGCATCAAGCGGATGTGCCATGCTTACCAGTTGCTGATAAAGGT 63
Qy 77 GlyLeuLysAspGluAsnIleIleValPheMetTyrAspIleAlaPheSerSerGlu 96
Db 64 CGAGTCAAGAAGAAACATTTGGTGTATGATGATGATATAGCTTATAACGCCATG 123
Qy 97 AsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyVal 116
Db 124 ATCCAGACCGGAGTCATCATCAACCATCTCCAGGGCCGACGTTATGCTGCTTCTGGTGA 183
Qy 117 ProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGly 136
Db 184 CTTAAGGATTACACCGGTGAGGAGGTAAACACCTCAGAACCTATATGCTGCTTCTTGGG 243
Qy 137 AsnGluSerGlyValThrGlyLysValLysValLysValLysValLysValLysValLys 156
Db 244 GACAAGAGTAAAGTTAAAGGTGGAAGTGGCAAGGTGATCAACCAATCCCGAGGATAGG 303
Qy 157 IlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAsp 176
Db 304 ATATTATTTCTACTCCGATCATGGAGGTCCCGAGTCTTGGGATGCCAAACGCCACCA 363
Qy 177 GluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyr 196
Db 364 TTCGTTTATGCCATGGATTTTATTGATGTTTGAAGAAGAAACATGCAAGTGGAGGTAC 423
Qy 197 AsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyLeuLeu 216
Db 424 AAGGAGATGGTTATATACATAGAAAGCTGTGAGAGTGGGAGCATATTTGAGGGTATATG 483
Qy 217 LysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGly 236
Db 484 CCCAAGGATCTGAATATTTATGTGACAACTCGCTCAATGCAAGAGACAGAGTTTGGGA 543
Qy 237 ValTyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAsp 256
Db 544 ACTTATTGTCCTGGGATGAATCTCTCTCCACAGAGATACGTAACTTGGCTGGGGAT 603
Qy 257 ThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeu 276
Db 604 TTATACAGCGTTTCTTGGATGGAAGATAGTGAGACTCACAACTCAAAAAGGAAACGGTA 663
Qy 277 GluGlnGlnTyrHisValValLysArgValGlySerAspValProGluThr----- 294
Db 664 CAACAGCAATACCAGTCGGTAAGGAAACGGACTTCAAAATCTCAACAGCTATAGGTTGGT 723

```

295 SerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIle 314
 724 TCTCATGTGATGCAATACGCTACACTACTACTGCTGAAACCTTACTTGTACCAT 783
 315 GlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerProIleSer 334
 784 GGTTTGATCTCCACCGCTGAACITTT-----CCCTCCACAC 819
 335 AsnSerGlyLeu-----ValAsnProArgAspIleProLeuLeuTyr 348
 820 AACGGCAACTAGAACCTAAATGGAAGTGTGTAACACAGAGAGATCGACAGCTTTTGTTC 879
 349 LeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLys 368
 880 ATGTGGCAAAATGATCATCAGAGATCAACACCATCAACCGGAAAGAAAGACTCACATCTGGAA 939
 369 LysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeu 388
 940 CAGATTACAGACAGACTGAAGCATAGGAATCATTTGGATGGCAGTGTGGAATTGATTGGA 999
 389 ArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrGly 408
 1000 GTTTTGTGTATGGACCGAGAAAGTTCTTCGGTTCTACATTCCGTGAGGCTCCTGT 1059
 409 GlnProLeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHis 428
 1060 CTGCCCTAGTGTGATGATGGACATGCTTGAATCTAGTGTAGAGTGTTCGAACAC 1119
 429 CysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsn 448
 1120 TGTGGGTCACTGACTCAGTATGGCATGAGAAACACATCGCGGCAATCGCGACGTTTGGCAAC 1179
 449 MetGlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCys 464
 1180 AGCGGGCTT-----TCTAAGCGCTTCCATGGAGAGGCTTGT 1215
 RESULT 14
 ID AAQ050564 standard; cDNA to mRNA; 1152 BP.
 AC AAQ050564;
 XX
 XX
 DT 24-MAY-1994 (first entry)
 XX
 XX Asparaginyl endopeptidase sequence (6).
 XX
 XX Asparaginyl endopeptidase; Canavalia ensiformis; seed;
 KW L-asparagine; primer; PCR; protein fragmentation;
 KW peptide synthesis; ss.
 XX
 XX Canavalia ensiformis.
 XX
 XX JP05276960-A.
 XX
 PD 26-OCT-1993.
 XX
 XX 07-AUG-1992; 92JP-0231602.
 XX
 PR 07-FEB-1992; 92JP-0056023.
 XX
 PA (SYUZU) SYUZU T.
 XX
 XX WPI; 1993-373587/47.
 XX
 XX New gene for encoding asparaginyl endo-peptidase - comprises 8
 PT specified DNA sequences
 XX
 XX Claim 1; Page 14-15; 35pp; Japanese.
 PS
 CC A gene encoding asparaginyl endopeptidase is claimed.
 CC 8 DNA sequences are given (AAQ050564-66). The enzyme is a protease
 CC derived from a seed of Canavalia ensiformis which selectively

CC hydrolyses C-terminus amide bond of L-asparagine residue
 CC (see AAQ0303 and AAQ0304).
 CC The enzyme is useful for protein fragmentation and enzymatic
 CC peptide synthesis.
 CC The primers given in AAQ050567-68, AAQ050576-77 and AAQ050583-90 were used
 CC in the isolation of the fragments given in AAQ050569-75 and AAQ050578-79,
 CC by PCR.
 XX
 SQ Sequence 1152 BP; 340 A; 210 C; 297 G; 305 T; 0 other;
 Alignment Scores:
 Pred. No.: 1,26e-87 Length: 1152
 Score: 930.00 Matches: 192
 Percent Similarity: 65.38% Conservative: 63
 Best Local Similarity: 49.23% Mismatches: 119
 Query Match: 40.34% Indels: 16
 DB: 14 Gaps: 3
 US-09-934-066-2 (1-466) x AAQ050564 (1-1152)
 QY 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
 DB 7 GGTACCCCGTGGCGGCTTCTCGTGGCTGTTCAACAGGTTATGCAATTTATAGCATCAA 66
 QY 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
 DB 67 GCGATGTGTGCTTACCGATTCCTGATTAAGGTGGAGTCAAGAAGAACAACAT 126
 QY 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
 DB 127 GTGGTGTATTATGATGATATAGCTTATACGCCATGAATCCAGACCCGAGTGCATC 186
 QY 104 IleAsnLysProAspGlyGlyValTyrLysGlyValProLysAspTyrThrLysGlu 123
 DB 187 ATCAACCATCTCCAGGGCCAGACGCTGATGCTGTGTTACCTAAGGATTACACCGGTGAG 246
 QY 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
 DB 247 GACGTAACACCTGAGAACCTATATGCTCTCTTGGGACAGAGTAAAGTTAAAGGT 306
 QY 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
 DB 307 GGAAGTGCGAAGTGTATCAACAGCAATCCGGAGGATAGGATATTTATATTCTACTCCGAT 366
 QY 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
 DB 367 CATGGAGTCCCGGAGTCTTGGGATGCCAACGCCACCATTCGTTACGCCATGGATTTT 426
 QY 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
 DB 427 ATTGATGTTTGAAGAAGAACAATCCAACTGAGGCTCAAGAGGATGGTTATATACATA 486
 QY 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
 DB 487 GAAGCTTGTGAGGTGGGAGCATATTTGAGGATATAATGCCCAAGGATCTGAATATTTAT 546
 QY 224 AlaValThrAlaAlaAsnSerLysGluSerTrpGlyValTyrCysProGluSerTyr 243
 DB 547 GTCACAACTCGCTCAAAATGCACAGAGAACAAGTTTGGAACTTATTCCTCTGGCATGAT 606
 QY 244 ProProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
 DB 607 CCTCCTCCACCAAGAGAGTACCTAACTTGCCTGGGGATTTTATACAGCGTTCTTTCGATG 666
 QY 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnIleValVal 283
 DB 667 GAAGATAGTACACTCACAATCTAAAGGCAACGCTACACCAACCAATACCAGTCCGTA 726
 QY 284 LysArgArgValGlySerAspValProGluThr-----SerHisValCysArgPheGly 301
 DB 727 AGCAACGCGACTTCAAAATCTTAACAGCTATAGTGTGTTCTCATGTGATGCAATACCGT 786
 QY 302 ThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAsp 321


```

Db 787 GACACTAATCTACTGCTGAAAGGCTTACTTGTACCATGGTTTGTGATCCTGCCACCGTG 846
Qy 322 AsnPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeu----- 338
Db 847 AACTTT-----CCTCCACACACGGCAACCTAGAGCTAAA 882
Qy 339 -----ValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLys 355
Db 883 ATGGAAGTTGTTAACACAGAGATGCCAGAGCTTTTGTTCATGTGCAAAATGATATCAGAGA 942
Qy 356 AlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsn 375
Db 943 TCACCAACCATCAACCGGAAAGAGACTCACATCTCGTGAACAGATTACAGACAGACTGAAG 1002
Qy 376 HisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThr 395
Db 1003 CATAGGAATCATTTGGATGGCAGTGTGGAATGTATGGAGTTTGTTCATATGGACAGGA 1062
Qy 396 AsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrp 415
Db 1063 AAAAGTTCTCGGTTCTACATTCCTGAGGGCTCTGCTGCTGCCCTAGTTGATGATTGG 1122
Qy 416 AspCysPheLysThrLeuValAsnSerPhe 425
Db 1123 ACATGCTTGAATCTATGTTAGAGTGTC 1152

```

RESULT 15

AAQ50565
ID AAQ50565 standard; cDNA to mRNA; 894 BP.

XX AC AAQ50565;
XX

DT 24-MAY-1994 (first entry)
XX

DE Asparaginylendopeptidase sequence (7).
XX

KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
KW L-asparagine; primer; PCR; protein fragmentation;
KW peptide synthesis; ss.
XX

OS Canavalia ensiformis.
XX

XX PN JP05276960-A.
XX

PD 26-OCT-1993.
XX

XX PF 07-AUG-1992; 92JP-0231602.
XX

PR 07-FEB-1992; 92JP-0056023.
XX

XX (SYUZ/) SYUZU T.
XX

XX PA WPI; 1993-373587/47.
XX

DR New gene for encoding asparaginyl endo-peptidase - comprises 8
XX specified DNA sequences
XX

PS Claim 1; Page 15; 35pp; Japanese.
XX

CC A gene encoding asparaginylendopeptidase is claimed.
CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
CC derived from a seed of Canavalia ensiformis which selectively
CC hydrolyses C-terminus amide bond of L-asparagine residue
CC (see AA43033 and AA43041).
CC The enzyme is useful for protein fragmentation and enzymatic
CC peptide synthesis.
CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
CC by PCR.
XX

XX SQ Sequence 894 BP; 265 A; 166 C; 229 G; 234 T; 0 other;

Alignment Scores:

Pred. No.: 4.77e-78 Length: 894
Score: 891.00 Matches: 162
Percent Similarity: 74.1% Conservative: 47
Best local Similarity: 57.45% Mismatches: 71
Query Match: 36.31% Indels: 2
DB: 14 Gaps: 1

US-09-934-066-2 (1-466) x AAQ50565 (1-894)

```

Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
Db 7 GGTACCCGGTGGGGGTTCTCGTGGTTCACACGGTTATGGAATATTAGGATCA 66
Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
Db 67 CGGATGTGTGCCATGCTTACAGTGTCTGATAAAGGTGGAGTCAAGAGAAAACATT 126
Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
Db 127 GTGGTGTATTATGATGATATAGCTTATAACGCCATGAATCCAGACCCGGAGTCATC 186
Qy 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
Db 187 ATCAACCATCTCTCAGGGCCAGACGTGTATGCTGGTACCTAAGGATTACACCGGTG 246
Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
Db 247 GACGTAAACACCTGAGAACCTATATGCTGTCTATCTTGGGGACAGAGTAAGATTAAAG 306
Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
Db 307 GGAAGTGGCAAGTGCATCAACAGCAATCCGGAGGATAGGATATTTATTTACTCCGAT 366
Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
Db 367 CATGGAGTCCCGGAGTTCTTGGGATGCCAACCCACCATTCGTTTACGCCATGATTTT 426
Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 427 ATTGATGTTTGAAGAAGAAACATGCAAGTGGAGGCTACAAGGAGATGGTTATATACATA 486
Qy 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
Db 487 GAAGCTTGTGAGAGTGGGAGCATATTTGAGGGTATATATGCCCAAGGATCTGAATATTTAT 546
Qy 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
Db 547 GTGACAACTGCTCAAAATGCACAGAGAACTTTTGGAACTTATTGTCTCTGGGATGAAT 606
Qy 244 ProProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
Db 607 CCTCTCCACAGAGAGTACGTACTTGCCTGGGGGATTTATACAGCGTTCTTGGGATG 666
Qy 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal 283
Db 667 GAAGATAGTGAGACTCACAAATCTAAAAGGGAACGGGTACACACAGTACACAGTCGCTA 726
Qy 284 LysArgArgValGlySerAspValProGluThr-----SerHisValCysArgPheGly 301
Db 727 AGGAACCGGACTTCAAAATCTTAACAGCTATAGTGTGTGGTTTGTCTCATGTGCAATACG 786
Qy 302 ThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAsp 321
Db 787 GACACTAACATTACTGCTGAAAAGCTTTTACTTGTACCATGGTTTGTGCTCTGCCACCG 846
Qy 322 AsnPhe 323
Db 847 AACTTT 852

```

Search completed: May 29, 2003, 18:27:15
Job time : 329 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 18:17:59 ; Search time 82 Seconds
(without alignments)
1742.822 Million cell updates/sec

Title: us-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFOLVFLHALLIF.....CNGMVVDVKQTVAIOACSM 466

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delep 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US0934066/runat_19052003_163700_24409/app_query.fasta_1.647
-DB-Issued_Patents_NA_QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blos62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=txt -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0934066.cgn_1_1_61@runat_19052003_163700_24409 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	828	33.7	1936	2	US-09-139-424-3
2	828	33.7	2030	3	US-08-706-216-3
3	744	30.3	1855	3	US-08-928-613-1
4	744	30.3	1855	3	US-09-193-524-1
5	178	7.3	220	2	US-08-928-613-15
6	178	7.3	220	3	US-09-193-524-15
7	168	6.8	219	2	US-08-928-613-11
8	168	6.8	219	3	US-09-193-524-11
9	137	5.6	1860	4	US-09-134-001C-1411
10	136	5.5	544	4	US-09-280-116-247
11	135.5	5.5	155	2	US-08-928-613-5
12	135.5	5.5	155	3	US-09-193-524-5

13	116	4.7	229	2	US-08-928-613-13	Sequence 13, Appl
14	116	4.7	229	3	US-09-193-524-13	Sequence 13, Appl
15	115.5	4.7	9510	4	US-09-453-702B-256	Sequence 256, App
16	114.5	4.7	331	2	US-08-928-613-10	Sequence 10, Appl
17	114.5	4.7	331	3	US-09-193-524-10	Sequence 10, Appl
18	114.5	4.7	1743	4	US-09-367-583-1	Sequence 1, Appl
19	112.5	4.6	4850	4	US-09-443-184-34	Sequence 34, Appl
20	104.5	4.3	206	2	US-08-928-613-22	Sequence 22, Appl
21	104.5	4.3	206	3	US-09-193-524-22	Sequence 22, Appl
22	103.5	4.2	199	2	US-08-928-613-18	Sequence 18, Appl
23	103.5	4.2	199	3	US-09-193-524-18	Sequence 18, Appl
24	103.5	4.2	13121	4	US-08-961-527-126	Sequence 126, App
c	103	4.2	4031	1	US-08-471-033-49	Sequence 49, Appl
25	103	4.2	4031	2	US-08-471-044-49	Sequence 49, Appl
26	103	4.2	4031	2	US-08-463-483A-49	Sequence 49, Appl
27	103	4.2	4031	2	US-08-471-046A-49	Sequence 49, Appl
28	103	4.2	4031	2	US-08-470-566B-49	Sequence 49, Appl
29	103	4.2	4031	2	US-08-469-334-49	Sequence 49, Appl
30	103	4.2	4031	3	US-09-300-529-49	Sequence 49, Appl
31	103	4.2	12687	1	US-08-676-169-1	Sequence 1, Appl
32	103	4.2	12687	3	US-08-981-459-1	Sequence 1, Appl
33	103	4.2	12687	4	US-09-063-431A-1	Sequence 1, Appl
34	103	4.2	12687	4	US-08-991-840A-2	Sequence 2, Appl
35	102.5	4.2	11464	4	US-09-454-721A-3	Sequence 3, Appl
36	100.5	4.1	11459	4	US-08-245-294-7	Sequence 7, Appl
37	100	4.1	1791	1	US-08-474-499-7	Sequence 7, Appl
38	100	4.1	1791	1	US-08-307-279A-7	Sequence 7, Appl
39	100	4.1	1791	4	US-09-525-310-7	Sequence 7, Appl
40	100	4.1	1791	4	PCT-US95-06211-7	Sequence 7, Appl
41	100	4.1	2877	4	US-09-619-353-1	Sequence 1, Appl
42	100	4.1	3792	2	US-08-992-334-1	Sequence 1, Appl
43	99.5	4.1	3792	3	US-08-302-752-1	Sequence 1, Appl
44	99.5	4.1	5234	2	US-08-992-334-2	Sequence 2, Appl
45	99.5	4.1	5234	2		

ALIGNMENTS

RESULT 1

US-09-139-424-3
; Sequence 3, Application US/09139424
; Patent No. 5985832
; GENERAL INFORMATION:
; APPLICANT: Roodman, G. David
; APPLICANT: Reddy, Sakamuri V.
; APPLICANT: Choi, Sun-Jin
; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE FOR
; TITLE OF INVENTION: OSTEOCLAST INHIBITORY FACTOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/139,424
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/772,441
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UTSK:295
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1936 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 136..1434

; US-09-139-424-3

Alignment Scores:

Pred. No.:	9,15e-91	Length:	1936
Score:	828.00	Matches:	179
Percent Similarity:	55.34%	Conservative:	80
Best Local Similarity:	38.25%	Mismatches:	157
Query Match:	33.74%	Indels:	52
DB:	2	Gaps:	13

US-09-934-066-2 (1-466) x US-09-139-424-3 (1-1936)

```

Qy      8 PheGlnIleuValPheLeuHisAlaLeuIlePheSerAlaGluSerArgLysThr 27
Db      142 TGAAGAGTACGTATCTCTCAGTGTGGCCCTGGGCATGTGGTGC-----186

Qy      28 GlnLeuLeuAsnAspAsnValGluSerSerLysSerAlaLysGlyThrArgTrp 47
Db      187 -----GTTCTATAGATGATCTCTGAGATGGAGGAGCAAGCACTGG 225

Qy      48 AlaValLeuValAlaGlySerAsnGluTyrAsnTyrArgHisGlnAlaAspIleCys 67
Db      226 GTGGTATGCTGTCAGGTTCAATGCTGATATATATATAGCACCAGGACGCGCTGC 285

Qy      68 HisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleValPheMet 87
Db      286 CATGCGCTACCATCATTCACCGCAATGGATTCCTCGACGAACAGATGCTGTGATGATG 345

Qy      88 TyrAspAspIleAlaPheSerGluAsnProArgProGlyValIleIleAsnLysPro 107
Db      346 TACGATGACATGCTTACTCTCAAGACAATCCCATCCAGCAATGTCATCAACAGGCC 405

Qy      108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
Db      406 AATGGCACAGATGCTATCAGGAGTCCCGAAGGACTACCTGGAGGAGGATTTACCCCA 465

Qy      128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly---GlyAsnGly 146
Db      466 CAAATTTCTCTGCTGTGTGAGCGGATCCAGAGCAGTGAAGGCGATAGGATCCGCG 525

Qy      147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla 166
Db      526 AAAGTCTGAAGAGTGGCCCCCAGGATCAGCGTGTTCATTACTTCACTGACCATGGATCT 585

Qy      167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
Db      586 ACTGGAATCTGTTTTTCTCC---AATGAAGATCTTCATGTAAGGAGCCTGAATGAGACC 642

Qy      187 LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
Db      643 ATCCATTACATGTACAAACACAAATGTACCGAAGATGGTGTCTACATTGAAGCCTGT 702

Qy      207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
Db      703 GAGTCTGGCTCCATGATGAACAC---CTGCGGGATAACATCAATGTTATGCAACTACT 759

Qy      227 AlaAlaSerLysLysSerTrpGlyValTyrCysProGluSerTyrProProPro 246
Db      760 GCTGCCAACCCAGAGAGTCTCTACCGCTGTACTATGATGAG-----804

Qy      247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
Db      805 -----AAGAGTCCACGTAACCTGGGGACTGGTACACGCTCAACTGGATGGAAGACTCG 858

```

```

Qy      267 AspLeuHisAspMetSerLysGluThrLeuGluGlnIntyRHisValValLysArgArg 286
Db      859 GACGTGGAAGATCTGACTAAAGAGACCTCGACACGAGTACCACTGGTAAATCCGAC 918

Qy      287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db      919 ACC-----AACACCAGCCACCTCATGCAGTATGCAACAAACAAATCTCC 963

Qy      307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
Db      964 ACCATCAAGTGCATCAGCTTTTCAGGT-----ATGAAA 996

Qy      327 GluSerPheSerSer-----ProIleSerAsnSerGlyLeuValAsnPro 341
Db      997 CGCAAAAGCCAGTCTCCCGTCCCTACCTCCAGTCACACACACTTGACCTACCCCGCAGC 1056

Qy      342 ArgAspIleProIleLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu 361
Db      1057 CCTGATGCTCCTCTCACCATCATGAAAGAACTGATGAACACC-----AATGATCTG 1110

Qy      362 ---GluSerLysGluAlaGlnLysLysLeuAspGluLysAsnHisArgLysGlnIle 380
Db      1111 GAGGAGTCCAGCAGCTCAGGAGAGATCCAGCGGATCTGGATGCGGACCTCATTT 1170

Qy      381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db      1171 GAGAAGTCAGTGCCTAAGATCCTCTCTGCTGGCAGCGTCCGAGGCTGAGGAGGAGCAG 1230

Qy      401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db      1231 CTCTCTGTCGACAGAGCC-----CCGCTCACGGGGAC---AGCTCTACCCAGAG 1278

Qy      421 LeuValAsnSerPheLysAsnHisCys-----GlyAlaThrValHisTyrGly 436
Db      1279 GCCTGCTGCACTTCCGGACCCACTGCTTCACTGGCAGCTCCCGCAGTACGAGTATGCG 1338

Qy      437 LeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThr 456
Db      1339 TTGAGACATTTGACGTGCTGCTCAACCTTTGTGAGAACCGTATCCGCTTCACAGGATA 1398

Qy      457 ValSerAlaIleGluGlnAlaCys 464
Db      1399 AATTTGCTCATGGACACGCTGTGC 1422

RESULT 2
US-08-706-216-3
; Sequence 3, Application US/08706216
; Patent No. 6140098
; GENERAL INFORMATION:
; APPLICANT: Balasubramanian, Sri Ram
; APPLICANT: Ford, John
; APPLICANT: Gorman, Daniel M.
; APPLICANT: Zurawski, Gerard
; TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706.216
; FILING DATE: 30-AUG-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-498-1200
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2030 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 183..1484.
US-08-706-216-3

Alignment Scores:
Pred. No.: 9,9e-91 Length: 2030
Score: 828.00 Matches: 179
Percent Similarity: 55.34% Conservative: 80
Best Local Similarity: 38.25% Mismatches: 157
Query Match: 33.74% Indels: 52
DB: 3 Gaps: 13

US-09-934-066-2 (1-466) x US-08-706-216-3 (1-2030)

QY 8 PheGlnLeuValPheLeuHisAlaLeuLeuPheSerAlaGluSerArgLysThr 27
DB 189 TGAAGTAGCTGTTATCTCTACGTGGCCCTGGGCATTTGGTGC 233
QY 28 GlnLeuLeuAsnAspValGluSerSerAlaLysGlyThrArgTrp 47
DB 234 -----GTTCTATAGATGATCCTGAAGTAGGAGCAAGCACTGG 272
QY 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
DB 273 GTGGTGATCGTCGAGGTTCAAAATGGCTGGTATATATAGGACCAGGACGCGGTGC 332
QY 68 HisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMet 87
DB 333 CATGCTACCATGATCATTCACCGCATGGATTCCTGACCAAGACAGATCGTTGTGATGATG 392
QY 88 TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
DB 393 TAGCATGACATGCTTACTCTGAAGACAATCCCACTCCAGGAATGTGTATCAACAGGCC 452
QY 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
DB 453 AATGGCACAGATGCTATCAGGAGTCCCGAAGGACTACACTGGAGAGATGTTACCCCA 512
QY 128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly---GlyAsnGly 146
DB 513 CAAATTTCTTCTGTTGAGGCGGATGCGAGACAGTGAAGGCGATAGATCGGCG 572
QY 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla 166
DB 573 AAAGTCTGAAGAGTGCCGCCAGGATCAGCTGTCTATTACTTCACTGACCATGGATCT 632
QY 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
DB 633 ACTGGAACTACTGTTTTCCTTCCTTCAATGAAGATCTTCAATGAAGGACCTGAATGAGACC 689
QY 187 LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
DB 690 ATCATTCATACATGTACAAACACAAATGTACCGAAAGATGGTGTCTACATTGAAGCTGT 749
QY 207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
DB 750 GAGTCTGGTGGTCCATGATGAACACAC---CTGCCGGATAACATCAATGTTTATGCAACTACT 806
QY 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProPro 246

```

RESULT 3

US-08-928-613-1

Sequence 1, Application US/08928613

Patent No. 5840562

GENERAL INFORMATION:

APPLICANT: Diep, Dinh

APPLICANT: Braxton, Scott M.

APPLICANT: Deleane, Angelo M.

TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESS: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/193,524
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/928,613
 ; FILING DATE:
 ; APPLICATION NUMBER: 08/567,506
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Luther, Barbara J.
 ; REGISTRATION NUMBER: 33954
 ; REFERENCE/DOCKET NUMBER: PF-0048 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-853-0555
 ; TELEFAX: 415-852-0195
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1855 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; IMMEDIATE SOURCE:
 ; LIBRARY: Adrenal
 ; CLONE: 100877

US-09-193-524-1

Alignment Scores:

Pred. No.: 1-51e-80 Length: 1855
 Score: 744.00 Matches: 186
 Percent Similarity: 54.63% Conservative: 70
 Best Local Similarity: 38.43% Mismatches: 128
 Query Match: 30.32% Indels: 68
 DB: 3 Gaps: 14

US-09-934-066-2 (1-466) x US-09-193-524-1 (1-1855)

QY 8 PheGlnIleValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr 27
 DB 60 TGGAAAGTAGTGTATCTCTCAGTGTGGCCCTGGGAATTGGTGCC----- 104
 QY 28 GlnLeuLeuAsnAspAsnValGluSerSerAspLysSerAlaLysGlyThrArgTrp 47
 DB 105 -----GTTCCTATAGATGATCCTGAAGATGGAGGCAAGCACTGG 143
 QY 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
 DB 144 GTGGTGATCGTGGCAGGTTCAATGCTGTGTATATATAGGCACCGACGACCGCTGC 203
 QY 68 HisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMet 87
 DB 204 CATGCTACCACTTCAATCCCGCAATGGATTCCTGCGCAACAGATCGTTGTGATTATG 263
 QY 88 TyrAspIleAlaPheSerGluAsnProArgProGlyValIleIleAsnLysPro 107
 DB 264 TACGATGACATAGCTTACTCTGAAGACAATCCCATCCAGGAATTGTGATCAACAGGCC 323
 QY 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
 DB 324 ANTGGCAGATGCTATCAGGAGTCCCGAGGACTACACTGGAGATGTTACCCCA 383
 QY 128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly---GlyAsnGly 146
 DB 384 CAAATTTCTCTGTGTGTGAGAGCGATGAGAGCAGTGAAGGTTATAGGATCCCGC 443
 QY 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla 166
 DB 444 AAGGTCCTGAAGAGTGGTCCCGCAGGATCAGCTGTTCATTATTTCTACCATCGATCT 503
 QY 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
 DB 504 TCTGGNATCTGGTTTCCCC---AATGAAGATCTCATGTAAGAGCCTGATTAGACC 560

QY 187 LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
 DB 561 ACCCATTCATTTTCAAAACAAATGTTACCGAAAGATGGTGTCTACATTAGGCGCTGT 620
 QY 207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
 DB 621 GAGTCTGGGTCCATGATGAACACAC---CTGCGGATAAATCAATGTTTATGCAACTACT 677
 QY 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProPro 246
 DB 678 GCTGCCAACCCAGAGAGTCTCTACGCCTGTACTATCATCAG----- 722
 QY 247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
 DB 723 -----AAGAGGTCCACGTACCTGGGGACTGTACAGCGTCAACTGGATGGAAGACTCG 776
 QY 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArg 286
 DB 777 GACGTGGAAGATCTGACTAAGAGACCCCTGCACAGCAGTACCACCTGGTAAATCGCAC 836
 QY 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
 DB 837 ACC-----AACACCACCGCAGTCATGCAGTATGGAACAAACAATCTCC 881
 QY 307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
 DB 882 ACCATGAAGTGTGATGTCAGTTTCAGGCT-----ATGAAA 914
 QY 327 GluSerPheSerSer-----ProIleSerAsnSerGlyLeuValAsnPro 341
 DB 915 CGCAAAGCCAGTCTCCCTCCCTCCCTACCTCCAGTCACACACTTGACCTCAGCCCGCAGC 974
 QY 342 ArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu 361
 DB 975 CCGTGTGCTCTCACCATCATGAAAGAAAGAACTGATGAACACC-----AATGATCTG 1028
 QY 362 GluSerLys-----GluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
 DB 1029 GAGGAGTCCAGCAGCTCAGGAGGAGATCCAGCGGTATCTGGATCCAGGCAC----- 1082
 QY 378 LysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnVal 397
 DB 1083 -----CTCATCCGAGGTGAGGTGGAGCAG----- 1106
 QY 398 LeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspTrpAspCys 417
 DB 1107 -----CTCTCTCCGAGAGGCC-----CCGCTCAGCGGGCAC---ACCTGC 1145
 QY 418 PheLysThrLeuValAsnSerPheLysAsnHisCys 429
 DB 1146 TACCCAGAGGTCTGTGTGTACTTCCGCGACCACTGC 1181

RESULT 5

US-08-928-613-15
 ; Sequence 15, Application US/08928613
 ; Patent No. 5840562
 ; GENERAL INFORMATION:
 ; APPLICANT: Diep, Dinh
 ; APPLICANT: Braxton, Scott M.
 ; APPLICANT: Deleane, Angelo M.
 ; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible


```

; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,613
; FILING DATE:
; CLASSIFICATION: 524
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: Bone Marrow
; CLONE: 104286
;
US-08-928-613-11

```

```

Alignment Scores:
Pred. No.:      8.04e-12      Length:      219
Score:          168.00      Matches:      34
Percent Similarity: 60.98%      Conservative: 16
Best Local Similarity: 41.46%      Mismatches: 20
Query Match:      6.85%      Indels:      12
DB:                2      Gaps:      2

```

US-09-934-066-2 (1-466) x US-08-928-613-11 (1-219)

```

Qy      222  IletyAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241
Db      1  GTTATGCAACTNCTGTCGCCAACCCAGAGAGTCGTCCTACGGCTGTNACTATGATGAG 60
Qy      242  SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261
Db      61  -----AAGAGGTCACCGTACCTGGGGGACTGGTACAGCGTCAAC 99
Qy      262  TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHis 281
Db      100  TGGATGGAAGACTCGGACGTGGAGATCTGACTAAAGAGACCTTGCACAGCAGTACCAC 159
Qy      282  ValValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPheGly 301
Db      160  CTGTAAATCGCACACC-----AACACCGCCACGTCATGCATATGGA 204
Qy      302  ThrGlu 303
Db      205  AACAAA 210

```

RESULT 8

US-09-193-524-11

; Sequence 11, Application US/09193524

```

; Patent No. 6007997
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Delegeane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/193,524
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,613
; FILING DATE:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: Bone Marrow
; CLONE: 104286
;
US-09-193-524-11

```

```

Alignment Scores:
Pred. No.:      8.04e-12      Length:      219
Score:          168.00      Matches:      34
Percent Similarity: 60.98%      Conservative: 16
Best Local Similarity: 41.46%      Mismatches: 20
Query Match:      6.85%      Indels:      12
DB:                3      Gaps:      2

```

US-09-934-066-2 (1-466) x US-09-193-524-11 (1-219)

```

Qy      222  IletyAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241
Db      1  GTTATGCAACTNCTGTCGCCAACCCAGAGAGTCGTCCTACGGCTGTNACTATGATGAG 60
Qy      242  SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261
Db      61  -----AAGAGGTCACCGTACCTGGGGGACTGGTACAGCGTCAAC 99
Qy      262  TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHis 281
Db      100  TGGATGGAAGACTCGGACGTGGAGATCTGACTAAAGAGACCTTGCACAGCAGTACCAC 159
Qy      282  ValValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPheGly 301
Db      160  CTGTAAATCGCACACC-----AACACCGCCACGTCATGCATATGGA 204

```

```
Oy 302 ThrGlu 303
Db 205 AACAAA 210

RESULT 9
US-09-134-001C-1411
: Sequence 1411, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 1411
: LENGTH: 1860
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1411

Alignment Scores:
Pred. No.: 1,78e-06 Length: 1860
Score: 137.00 Matches: 104
Percent Similarity: 36.57% Conservative: 73
Best Local Similarity: 21.49% Mismatches: 191
Query Match: 5.58% Indels: 116
DB: 4 Gaps: 22

US-09-934-066-2 (1-466) x US-09-134-001C-1411 (1-1860)
Oy 22 AlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAspLysSer 41
Db 208 GCTGAATACGTAAA-----GTTGCAGATAAATAGTAGGAGGATGGTACGTAGGA 258
Oy 42 AlaLysGlyThrArgTirpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArg 61
Db 259 ATTGGTCATACACGTTGGGCG-----ACACATGGTGTCCCAAAATATGAAATTCACAT 312
Oy 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp--- 80
Db 313 CCACACGAGTCACATCTGCAGCTTTTACATTTAGTTCATATGTTGTAATGAAATTTAT 372
Oy 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 373 GAGAAATTAAGCTGAATATTATCTGATGTCATCTTTTTCATCAGAACT----- 423
Oy 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 424 -----GATACGGAAGTTATTGTACAAATAGTA-----GATTAT 456
Oy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 457 -----TTTTCTAGACAAGGATTAGCTACAGAAGATGCA 489
Oy 141 ValThrGlyGlyAsnGlyLysValValLys----- 150
Db 490 TTTACA-----AAAGTAGTTAAATATTATACATGGTTCATATGCTTTAGGATTA 537
Oy 151 SerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIle 170
Db 538 TTAGATGATTAATGATAAGATACATCTTTTATGTTGGCTAAAAACAAGCTCTCCGCTTTAGTA 597
Oy 171 AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValIleGlyMet 190
Db 598 GGTGTAGGTGAAGTTTCAATGTTTCTGCTGATGCTAGCAATGTTACAACTACA 657
Oy 191 HisLysArgLysLys-----TyrAsnLysMetValIleTyrValGluAlaCysGlu 207
```

```
Db 658 AACCAATACAAAGAGATACATCACCATGAATAGTATTATTGTTAAGCGAGACAGACAGTAGAA 717
Oy 208 SerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAla 227
Db 718 ATTAAAGATCTTGGAGGGGACATTTCAACACGCTGATAGCTATACGGCAGAAATAGATGCT 777
Oy 228 AlaAsnSerLysGluSerSerTirpGlyValTyrCysProGluSerTyrProProProPro 247
Db 778 GCTGATGCACAAAAGCGCTATATCATCATTAATGTTTAAAGAAATTCATGAACAGCGCT 837
Oy 248 Ser-----GluIleGlyThrCysLeuGlyAsp 256
Db 838 GCAGTGATGCGTCGCATTATTCAAGAAATATCAAGATGAAAGAGCTAATTTAAAAATCGAT 897
Oy 257 ThrPheSerIleSerTirpLeuGluAspSerAsp---LeuHisAspMetSerLysGluThr 275
Db 898 TCAGAGATTATTAAATGATGTAGCAGATGCTGATATTTACATCTGTTGCAGCTGCTACT 957
Oy 276 LeuGluGlnGlnThrHisVal-----ValLysArgArgValGly 288
Db 958 -----AGTTATCATCTCGATTGGTTGCTAAAGAAATTTATTGAAAAATGSCAGGT 1008
Oy 289 SerAspValProGluThrSerHisValCysArg-----PheGlyThrGluLysMet 305
Db 1009 -----CTACCTACTGAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1062
Oy 306 LeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGlu----- 319
Db 1063 TCTGAAAAACCACTATTATTATTTTACCAANTCTGTTGAAACAGCTGATAGTCGTGCT 1122
Oy 320 -----AsnAspAsnPheThrPheThrGluSerPheSer 330
Db 1123 GTATTAGTTGAAACAAATAAGTTAGGTCACAAATCATTAACAATTTACTAATGTTGCTGGT 1182
Oy 331 SerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGln 350
Db 1183 TCAACATTTATCAGCTGAAGCG-----GATCATACATTACTTTTACATGCTGGA 1230
Oy 351 ArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu 370
Db 1231 CTTGAGATTGCGTCGCATCTACAAAAGCATATACAGCGCAAAATGCTGTTTATCTATC 1290
Oy 371 LeuAspGlu-----LysAsnHisArgLysGlnIleAspGlnSerIleThrAspIle 387
Db 1291 TTATCTCAAATGTTGCTTAAATCATGCTGCTGAAACCGAT-----CTTGATTTA 1341
Oy 388 LeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThr 407
Db 1342 TTAAGAACTAGCTAAGGTTTACTACAGCTATT----- 1374
Oy 408 GlyGlnProLeuValAspAsp-----TrpAspCysPheLys 419
Db 1375 ---GAAACAATTTGTTGACGATGCACCTAAGATGGAGCAAAATGCAACGGATTCTTAAAA 1431
Oy 420 ThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyr 439
Db 1432 ACTACTCGTAATGCATTC-----TTCAATGGACCAACAATGATTATAATGTTAGTTTA 1485
Oy 440 ThrGlyAlaLeu 443
Db 1486 GAAGGTGCATTA 1497

RESULT 10
US-09-280-116-247
: Sequence 247, Application US/09280116A
: Patent No. 6331427
: GENERAL INFORMATION:
: APPLICANT: Robison, Keith E.
: TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
: FILE REFERENCE: 3800-24, 035800/176965
: CURRENT APPLICATION NUMBER: US/09/280,116A
: CURRENT FILING DATE: 1999-03-26
```

```
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 247
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: hemoglobinase
; NAME/KEY: misc.feature
; LOCATION: (1)..(544)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-247

Alignment Scores:
Pred. No.: 2.98e-07 Length: 544
Score: 136.00 Matches: 25
Percent Similarity: 70.43% Conservatives: 6
Best Local Similarity: 56.82% Mismatches: 13
Query Match: 5.54% Indels: 0
DB: 4 Gaps: 0

US-09-934-066-2 (1-466) x US-09-280-116-247 (1-544)
QY 53 GlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIle 72
Db 34 GGATCTAAATCTACCAAAATCATTTCTCCAGGAGACGCGTGCCTACCAAGATC 93
QY 73 LeuArgLysGlyGlyLeuLysaspGluAsnIleIleValPheMetTyrAspIleAla 92
Db 94 ATTCAACCGCAATGGGATTCCTCAGCAAGACAGATCGTTGTGATGATGATGATGATGAT 153
QY 93 PheSerSerGlu 96
Db 154 TACTCTGAAGAG 165

RESULT 11
US-08-928-613-5
; Sequence 5, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Delegeane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/928,613
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: 08/567,506
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: THP-1
; CLONE: 075848
US-08-928-613-5

Alignment Scores:
Pred. No.: 4.15e-08 Length: 155
Score: 135.50 Matches: 25
Percent Similarity: 69.23% Conservatives: 11
Best Local Similarity: 48.08% Mismatches: 11
Query Match: 5.52% Indels: 5
DB: 2 Gaps: 1

US-09-934-066-2 (1-466) x US-08-928-613-5 (1-155)
QY 252 ThrCysLeuGlyAspThrPheSerIleSerTyrLeuGluAspSerAspLeuHisAspMet 271
Db 13 AGTACTCTGGGACCTGGTACAGCTCACTCGATGGAGAACTCGGACCTGGAGATCTG 72
QY 272 SerLysGluThrLeuGluGlnGlnTyrHisValValLysArgValGlySerAspVal 291
Db 73 ACTAAGACACCTCGCACAGCAGTACCACCTGGTAAATCGCACACC----- 120
QY 292 ProGluThrSerHisValCysArgPheGlyThrGlu 303
Db 121 ---AACACCAAGCCAGCTCATGTCAGTATGGAACAAA 153

RESULT 12
US-09-193-524-5
; Sequence 5, Application US/09193524
; Patent No. 6007997
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Delegeane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/193,524
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,613
; FILING DATE:
; APPLICATION NUMBER: 08/567,506
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: THP-1
; CLONE: 075848
US-09-193-524-5

Alignment Scores:
Pred. No.: 4,15e-08 Length: 155
Score: 135.50 Matches: 25
Percent Similarity: 69.23% Conservative: 11
Best Local Similarity: 48.08% Mismatches: 11
Query Match: 5.52% Indels: 5
DB: 3 Gaps: 1

US-09-934-066-2 (1-466) x US-09-193-524-5 (1-155)
QY 252 ThrCysLeuGluYAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMet 271
Db 13 ACGTACCTGGGGGACTGTGACAGCAGTCACTGATGATGGAAGACTCGGACGTGGGAAGATCTG 72
QY 272 SerLysGluThrLeuGluGlnClyThrHisValLysArgValGlySerAspVal 291
Db 73 ACTAAGAGACCTCCACCAAGCAGTACCACCTGTAAATCGCACACC-----120
QY 292 ProGluThrSerHisValCysArgPheGlyThrGlu 303
Db 121 ---ACACCGACCGCTGATGATGATGGAACAAA 153

RESULT 13
US-08-928-613-13
; Sequence 13, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleageane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928.613
; FILING DATE:
; CLASSIFICATION: 524
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: Lung
; CLONE: 125569
US-08-928-613-13

Alignment Scores:
Pred. No.: 1,92e-05 Length: 229
Score: 116.00 Matches: 31
Percent Similarity: 56.76% Conservative: 11
Best Local Similarity: 41.89% Mismatches: 22
Query Match: 4.73% Indels: 10
DB: 2 Gaps: 2

US-09-934-066-2 (1-466) x US-08-928-613-13 (1-229)
QY 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
Db 5 TACATTGGAGCCTCTGAGTCTGGCTCCATGATGAACCAC---CTNCCGGATAACATCAAT 61
QY 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241
Db 62 GTTATGCAACTACTGCTGCCAACCCAGAGAGTCTCTACCCCTGTTACTATGATGAG 121
QY 242 SerTyrProProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle-5e 261
Db 122 -----AAGAGNCCACGTCCTGGGGGAGCTNGTACAAANGTNAAA 160
QY 261 rTrpLeuGluAsp-SerAspLeuHisAspMetSerLys 273
Db 161 NTNGATGGAAGAATTACAGACGAGGAAGATCTNNCTAAA 198

RESULT 14
US-09-193-524-13
; Sequence 13, Application US/09193524
; Patent No. 6007997
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleageane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/193.524
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928.613
; FILING DATE:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 229 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 IMMEDIATE SOURCE:
 LIBRARY: Lung
 CLONE: 125569
 US-09-193-524-13

Alignment Scores:
 Pred. No.: 1.92e-05 Length: 229
 Score: 116.00 Matches: 31
 Percent Similarity: 56.76% Conservative: 11
 Best Local Similarity: 41.89% Mismatches: 22
 Query Match: 4.73% Indels: 10
 Gaps: 2

US-09-934-066-2 (1-466) x US-09-193-524-13 (1-229)

Qy 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsh 221
 Db 5 TACATTGANGCCTGTGAGTCTGGTCCATGATGAACAC---CTNCCGGATAACATCAAT 61
 Qy 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerTrpGlyValTyrCysProGlu 241
 Db 62 GTTATGCAACTACTGCTGCCAACCACGAGAGTCTCTACGCTGTACTATGATGAG 121
 Qy 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle-Se 261
 Db 122 -----AAGAGNCCACGTACCTGGGGGACTNGTACAAAGTNAAA 160
 Qy 261 rTrpLeuGluAsp-SerAspLeuHisAspMetSerLys 273
 Db 161 NTNGATGGAAGATTACAGCAGGAGATCTNNCTAAA 198

RESULT 15

US-09-453-702B-256
 : Sequence 256, Application US/09453702B
 : Patent No. 6365723
 : GENERAL INFORMATION:
 : APPLICANT: Blattner, Frederick R.
 : Burland, Valerie
 : Perna, Nicole T.
 : Plunkett, Guy
 : Welch, Rod
 : TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
 : NUMBER OF SEQUENCES: 265
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Quarles & Brady
 : STREET: 1 South Pinckney Street
 : CITY: Madison
 : STATE: WI
 : COUNTRY: US
 : ZIP: 53701-2113

COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Word Perfect 8.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/453.702B
 : FILING DATE: 03-Dec-1999
 : CLASSIFICATION: <Unknown>
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 60/110,955
 : FILING DATE: 04-DEC-1998
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Seay, Nicholas J.
 : REGISTRATION NUMBER: 27386
 : REFERENCE/DOCKET NUMBER: 960296.95017
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166
 : INFORMATION FOR SEQ ID NO: 256:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 9510 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: double
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : SEQUENCE DESCRIPTION: SEQ ID NO: 256:
 US-09-453-702B-256

Alignment Scores:
 Pred. No.: 0.0116 Length: 9510
 Score: 115.50 Matches: 115
 Percent Similarity: 36.46% Conservative: 83
 Best Local Similarity: 21.18% Mismatches: 188
 Query Match: 4.71% Indels: 157
 Gaps: 30

US-09-934-066-2 (1-466) x US-09-453-702B-256 (1-9510)

Qy 18 LeuIlePheSerAlaGluSerArgLysThrClnLeuLeuAsnAspAsn----- 33
 Db 7486 ATTATTATGTAATCATGGGCAAAATACACCTTCCTCAATTCAGGAAGTAAACATATA 7545
 Qy 34 AspValGluSerSerAspLysSerAlaLys-----GlyThrArgTirAlaVal 49
 Db 7546 GATATTACATCTCTTGATAAAATAGTCATTATCATCTTTTGACTTCGCTAATATTATT 7605
 Qy 50 LeuValAlaGlySerAsnGlu---TyrTyrAsnTyrArgHisGlnAlaAspIleCysHis 68
 Db 7606 TTTGAACATTACACCAATGAGATATTTATAAGTTTGTGATAATCAAGGAGGATTT----- 7659
 Qy 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
 Db 7660 GTTATATCAATATTAAATAATGTCAGAGTGAAGCGGCAACATTAATGATTTCGGAATA 7719
 Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
 Db 7720 AACATGACTCTTTAGATAGTAGT-----GGTTCATTGATATATTTCCCTAGC 7767
 Qy 109 GlyGlu-----AspValTyrLysGlyValProLysAspTyrTyrLysGlu 123
 Db 7768 GGTGATATTTATCATATTTTCAGATATCTATAAA---ATGTCAGAGGTAGAAAAATCGTTT 7824
 Qy 124 AlaValAsnValGln-----AsnPheTyrAsnValLeuLeuGlyAsn 137
 Db 7825 AAATTAATGTAGAAAGAAAGAACCTGCATGATGATATCATCAATGTGGCTATTTTAA 7884
 Qy 138 GluSerGlyValThrGlyGlyAsnGlyLysValLysSerGlyProAsn-----Asp 155
 Db 7885 ACTTCATTATCTACAA-----ATAAAAAATCCCAATTAATGATGAT 7926
 Qy 156 AsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
 Db 7927 AGTGATTATATTTTATGTTTAGACAAC-----CCTAACCTCTCATCT----- 7968
 Qy 176 AspGluValMetAlaLysAspPheAsnGluVal----- 186
 Db 7969 -----TACACATTGAACCTTTAATGACTGTCTGGATACATATCAAGTTTATGGAT 8019
 Qy 187 -----LeuGluLysMetHisLys-----Arg 193
 Db 8020 AATATTAGGGGTAGTTTTTACCCTTTTCATAAAAAATACTGTAAACATCGCCCTTAATGAA 8079
 Qy 194 LysLysTyrAsnLysMetValIleTyrValIcluaAlaCysGluSerGlySerMetPheGlu 213
 Db 8080 AAAAAATATATTTCCCTAAATCGGACTTTGATAAGTTGTCAATCAATATTGATGTTTGA 8139
 Qy 214 GlyIleLeuLys---LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGlu 232
 Db 8140 CAAGCATTAGAGGTAAAAATAAGAAATAGTTACAGAATAAGT-----AAA 8184

Qy	233	SerSerTrpGlyValTyr	-----CysProGlu-----	241
Db	8185	TTTACATGGGAACGTATGTTGATATCATCACC	TTTACATGAGAATTTCTCATCTG	8244
Qy	242	-----SerTyrProProProSerGluIleClyThrCysLeuGlyAsp	256	
Db	8245	GAGTTAGATGGGTTTAATTAATTTTTCACAACCC	--GAGTTAGATAGTCCCAATCTGCAC	8301
Qy	257	ThrPheSerIleSerTrp	-----LeuGluAspSerAspLeuHis	269
Db	8302	TCATTTTTCATATTTATATGATAATTTTCAGATCGTTGATAGTGCATATTAAGCTT	8361	
Qy	270	---AspMetSerLysGluThrLeuGluGlnGlnTyrHis	-----ValValLysArgArg	286
Db	8362	CTTTCATCTCAACAGAGAACAACAGCAACACCATCGTATTATTCTTTAAAGATAC	8421	
Qy	287	ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu	306	
Db	8422	TTTATAGACTCATTTGCTAAACACAGTATCATCTGACAGA	-----GAACAAATATA	8472
Qy	307	-----LysAspTyrLeuSerSerTyrIleGlyArgAsn	317	
Db	8473	TATCCTGTGATTTGTGATTCACCTGATCACTTCACAAGCGACATATATAGGCATCCCATTT	8532	
Qy	318	-----ProGluAsnAspAsnPheThrPheThrGlu	327	
Db	8533	AGAATAGTCTCGGGAATAAGACATTTGTATCCCATCAGAAGAAGACTTGTAATAATTCATAAGC	8592	
Qy	328	SerPheSerSerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeu	347	
Db	8593	AGCTCAAAAGAATATTTAAAGCAATATGGATGTGATTAAACNAATGTTATTGTTCTCT	8646	
Qy	348	TyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGln	367	
Db	8647	-----CAA	8649	
Qy	368	LysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIle	387	
Db	8650	AAAAACAACAAAGAAAAATATTCATAGTTAGTTTAAATAGCAATATCAAAANATCAT	8709	
Qy	388	LeuArgLeuSerValLysGlnThrAsnVal	-----LeuAsn-----	399
Db	8710	ATTGTATTGTCGGGAGTAATGACCGGNACATCAAGATTTTTCATCTAAATAATTCAGGT	8769	
Qy	400	-----LeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspTrpAspCys	417	
Db	8770	GATTATTGTTACAACTCAAAACATCATGGTGGGGTGTG	-----GTTATT	8820
Qy	418	PheLysThrLeuValAsnSerPhe	-----LysAsnHisCysGlyAlaThrValHisTyr	435
Db	8821	TTTAAAGATTTTATAAATACTGGTGGAAATATAATCTTACTTTAATCAACAGTTCCTTATT	8880	
Qy	436	GlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGln	455	
Db	8881	GATAATAAATTTGTCAGACAAATAGATAAATATATAACCCCAATGGGATTAAGATACAGAA	8940	
Qy	456	ThrValSer	458	
Db	8941	ACAGTCAGT	8949	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 19:19:39 ; Search time 186 Seconds
(without alignments)
3378.162 Million cell updates/sec

Title: us-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHQILVFLHALLIF.....CNMGVDVKQTVSAIEQACSM 466

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 845702 seqs, 674182571 residues

Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=slp
-O=/cgn2_1/USPTO.spool/US09934066/runat_19052003_163702_24460/app_query.fasta_1.647
-DB=PublishedApplications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09934066 -CGN_1_1_80 -runat_19052003_163702_24460
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT
-DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAPEXT=0.5 -FPGAPOP=6 -FPGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PublishedApplications_NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2454	100.0	1560	10	US-09-934-066-1 Sequence 1, Appli
2	2188	89.2	4320	10	US-09-934-066-3 Sequence 3, Appli
3	1246	50.8	1485	9	US-09-938-842A-1485 Sequence 1485, Ap
4	1160	47.3	1461	9	US-09-938-842A-927 Sequence 927, App

5	825	33.6	1974	9	US-10-208-408-18	Sequence 18, Appli
6	818	33.3	2002	12	US-10-002-600-10	Sequence 10, Appli
7	744	30.3	1855	9	US-09-967-796-1	Sequence 1, Appli
8	487	19.8	960	9	US-10-043-487-97	Sequence 97, Appli
9	450	18.3	405	10	US-09-878-574-4049	Sequence 4049, Ap
10	441	18.0	393	10	US-09-878-574-3641	Sequence 3641, Ap
11	412.5	16.8	392	10	US-09-878-574-3407	Sequence 3407, Ap
12	333	13.6	415	10	US-09-960-352-13205	Sequence 13205, A
13	333	13.6	422	10	US-09-960-352-11028	Sequence 11028, A
14	333	13.6	426	10	US-09-960-352-1596	Sequence 1596, Ap
15	307	12.5	407	10	US-09-960-352-5110	Sequence 5110, Ap
16	289	11.8	415	10	US-09-960-352-10118	Sequence 10118, A
17	245.5	10.0	282	10	US-09-878-574-7312	Sequence 7312, Ap
18	241	9.8	455	9	US-09-918-995-8935	Sequence 8935, Ap
19	226.5	9.2	375	10	US-09-878-574-2361	Sequence 2361, Ap
20	210	8.6	184	10	US-09-960-352-11538	Sequence 11538, A
21	190.5	7.8	459	9	US-09-918-995-11442	Sequence 11442, A
22	178	7.3	220	9	US-09-967-796-15	Sequence 15, Appli
23	168	7.3	367	10	US-09-878-574-1704	Sequence 1704, Ap
24	168	6.8	219	9	US-09-967-796-11	Sequence 11, Appli
25	157	6.4	454	10	US-09-878-574-4350	Sequence 4350, Ap
26	151	6.2	273	10	US-09-923-876-4813	Sequence 4813, Ap
27	146	5.9	15061	9	US-10-092-154-991	Sequence 991, App
28	146	5.9	15061	9	US-09-764-891-6914	Sequence 6914, Ap
29	146	5.9	15061	10	US-09-764-847-991	Sequence 991, App
30	144.5	5.9	442	10	US-09-920-300A-729	Sequence 729, App
31	144.5	5.9	442	12	US-10-033-528-729	Sequence 729, App
32	135.5	5.5	155	9	US-09-967-796-5	Sequence 5, Appli
33	133.5	5.4	280	10	US-09-878-574-5529	Sequence 5529, Ap
34	127	5.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
35	116	4.7	229	9	US-09-967-796-13	Sequence 13, Appli
36	115.5	4.7	9510	9	US-10-114-170-256	Sequence 256, App
37	114.5	4.7	331	9	US-09-967-796-10	Sequence 10, Appli
38	114.5	4.7	1743	9	US-10-202-211-1	Sequence 1, Appli
39	108.5	4.4	1074	10	US-09-974-300-2794	Sequence 2794, Ap
40	107	4.4	2988	9	US-09-738-626-347	Sequence 347, App
41	104.5	4.3	206	9	US-09-967-796-22	Sequence 22, Appli
42	104.5	4.3	244	10	US-09-878-574-10299	Sequence 10299, A
43	104	4.2	4959	9	US-10-025-950-1	Sequence 1, Appli
44	103.5	4.2	199	9	US-09-967-796-18	Sequence 18, Appli
45	103.5	4.2	12379	10	US-09-991-258-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1
US-09-934-066-1
; Sequence 1, Application US/09934066
; Patent No. US20020108149A1
; GENERAL INFORMATION:
; APPLICANT: Gruis, Darren B.
; APPLICANT: Jung, Rudolf
; TITLE OF INVENTION: Methods of Increasing Polypeptide Accumulation in Plants
; FILE REFERENCE: 35718/237251
; CURRENT APPLICATION NUMBER: US/09/934,066
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,804
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-934-066-1
Alignment Scores:
Pred. No.: 5.3e-279 Length: 1560
Score: 2454.00 Matches: 466
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0

DB	1060	CGCGCGGATATCTCTGCTATACCTCCAGAGAAAGATTCAAAAGCTCCAAATGGGATCA	1119
Qy	361	LeuGluSerLysGIuAlaGlnLysLysLeuLeuAspGluLysAsnHIsArGLysGlnIle	380
Db	1120	CTTGAAGCAAGAGCTCAGAGAAATTCGTTACCAAAAAGAAATCATAGCAAAACAATC	1179
Qy	381	AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu	400
Db	1180	GATCAGAGCATTCACAGACATTCCTCGGCTTTCAGTTAAACAAACCAATGCTTTAAATCTC	1239
Qy	401	LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspThrAspCysPheLysThr	420
Db	1240	TTAACTTCCACAACAACAGGACGCTCTCTAGACGATTTGGGATTCGTTCAAGACT	1299
Qy	421	LeuValAsnSerPheLysAsnHIsCysGlyAlaThrValHISLysrGlyLeuLysThr	440
Db	1300	CTAGTTAATAGCTTCAGANTCACTCGGTGCACGGTGCATTACGGATTGAAGTATACA	1359
Qy	441	GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle	460
Db	1360	GGAGCGCTTGCAATATCTGCATATGGGAGTGGATGGAAGCAAACTGTTTCAGGCATT	1419
Qy	461	GluGlnAlaCysSerMet 466	
Db	1420	GAACAAGCTTGTTCGATG 1437	
RESULT 2			
US-09-934-066-3			
; Sequence 3, Application US/09934066			
; Patent No. US20020108149A1			
; GENERAL INFORMATION:			
; APPLICANT: Gruis, Darren B.			
; TITLE OF INVENTION: Methods of Increasing Polypeptide			
; TITLE OF INVENTION: Accumulation in Plants			
; FILE REFERENCE: 35718/237251			
; CURRENT APPLICATION NUMBER: US/09/934.066			
; CURRENT FILING DATE: 2001-08-21			
; PRIOR APPLICATION NUMBER: US 60/226,804			
; PRIOR FILING DATE: 2000-08-21			
; NUMBER OF SEQ ID NOS: 3			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 3			
; LENGTH: 4320			
; TYPE: DNA			
; ORGANISM: Arabidopsis thaliana			
US-09-934-066-3			
Alignment Scores:			
Pred. No.:	5.91e-247	Length:	4320
Score:	2186.00	Matches:	463
Percent Similarity:	67.89%	Conservative:	0
Best Local Similarity:	67.89%	Mismatches:	3
Query Match:	89.16%	Indels:	219
DB:	10	Gaps:	7
US-09-934-066-2 (1-466) x US-09-934-066-3 (1-4320)			
Qy	1	MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe	20
Db	2020	ATGCTAGTCTCTTGCTACTTTCAGATTCTGTTTTCTTCATGCTTTGCTTATATCTTC	2079
Qy	21	SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnAspValGluSerSerAspLys	40
Db	2080	TCAGCTCAGTCCCGCAAAACCCCAATTCCTCAACGATAATGATCTGTAATCTACGCCAAG	2139
Qy	41	SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTrpTyTrAsnTyr	60
Db	2140	AGTGC AAAAGGCACACCATGGCTGTTTTAGTTGCTGGATCAAAATCAATATATATACTAC	2199
Qy	61	ArgHis	62

Db 2200 AGGCATCAGGTGTTAAATTAATGTTGAACGTTTAAACATAACAAAAAGGTCCA 2259
Qy 63 -----GlnAlaAspIleCysHis 68
Db 2260 AGCGAGATTTCTATGAACATAACACCGAGGTTTATTTACACAGGCTGACATATGCCAC 2319
Qy 69 AlatyrglnIleLeuArgLysGlyLeuLysaspGluAsnIleIleValPheMetTyr 88
Db 2320 CGGTATCAGATPACTCCGAAAGGCGGTTTAAAGATGAAACATCATTTGTATGAT 2379
Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db 2380 GATGATATCGCGTTTCTCGAGAACTTCTAGGCTGGAGTTATCATTAATAAACAGAT 2439
Qy 109 GlyGluAspValTyrLysGlyValProLys ----- 118
Db 2440 GGAGAAGATGTTTATAAGGAGTTCCCTAA - GGTTCTTATTCTACTTCTTTTGTGCGTTA 2498
Qy 118 ----- 118
Db 2499 TTTCTAGCTTGAATTCATATATATATATCAAGTTTGTGTTATTGTTGGGTAG 2558
Qy 119 AspTyrThrLysGluAlaValAsnValGlnAsnPhetYrasnValLeuLeuGlyAsnGlu 138
Db 2559 GACTACACTAAGAAGCTGTTAATGTTCAAACTTCTACAAATGTTTACTTTGGAATGAA 2618
Qy 139 SerGlyValThrGlyGlyAsnGlyLysValLysSerGlyProAsnAspAsnIlePhe 158
Db 2619 AGTGGCTGCACAGAGGAAATGGCAAGTTGTGAAAGTGTCTTAATGATAATATCTTC 2678
Qy 159 IleTyrThrAlaAspHisGlyAlaProGlyLeuIle ----- 170
Db 2679 ATCTATTATGCTGACCATGGAGCTCCTCGCTTAATAGGTTTCTTAATTTTATGAAATTA 2738
Qy 170 ----- 170
Db 2739 TTAGCTACCATCAATCCATATCTATAATAAAGATTTTCTCTGATACTAGCAACCGCG 2798
Qy 171 -----AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeu 187
Db 2799 ATTTTCTCAGCGATGCCACTGGTGATGAAGTTATGGCAAAAGATTTCAATGAAGTCTTG 2858
Qy 188 GluLysMetHisLysArgLysLysTyrAsnLysMetVal ----- 200
Db 2859 GAGAAGATGCATAAGAGAAAAATACAAAGATGTT - ATATAACTCAACCATTCGTTA 2917
Qy 200 ----- 200
Db 2918 CCTAGCTTTATACATATGTGTTCTGTTTGAATCTCTATGGTGTGTTTTTGGATGTT 2977
Qy 201 -----IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLys 218
Db 2978 TAGGTGATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAA 3037
Qy 219 AsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyr 238
Db 3038 AATCTCAACATATACGAGTGACTGCTGCTAATCTAAAGAGACAGCTGGGGAGTTTAC 3097
Qy 239 CysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPhe 258
Db 3098 TGTCTGAGTCATATCT 3157
Qy 259 SerIleSerTrpLeuGlu ----- 264
Db 3158 AGCATCTCTTGGCTTGAGGACAGCTACTGCAACAAAAAGATTCATCTTATGGACTA 3217
Qy 264 ----- 264
Db 3218 TTCGAATGATTTGTTCTTGAAGAATATTTGTTCAATTTGTTCTATGTTTGTGTGT 3277
Qy 265 -----AspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHisVa 282
Db 3278 GTTTGGGACAGTGACCTTCATGACATGAGCAAGAGACTTTGGAGCAACAATACACGTT 3337

Qy 282 lValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPheGlyTh 302
Db 3338 TGTAAAGAGAAGACTAGTAGTGTACCAGACATCTCTCATGTATGCCGTTTCGGAAC 3397
Qy 302 rGluLysMetLeuLysaspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAs 322
Db 3398 AGAGAAGATGCTTAAAGATTTATCTTCTCTTACATTTGGAAGAAATCTGTAACACGATAA 3457
Qy 322 nPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsnProAr 342
Db 3458 CTTTCACITTCACGGAATCTTTCTCTCACCNAATCTCTAATCTCGCTTGGTCAATCCGCG 3517
Qy 342 gAspIleProLeuLeuTyrLeuGlnArgLysIle ----- 353
Db 3518 CGATATTCTCTGTATACCTCCAGAGAAAGGT - GAGCTTTTTCGGGTTTTTTTGATCAT 3576
Qy 354 -----GlnLysAl 356
Db 3577 TTTAAACGAAAGACTTTTTCAGCATGTTTAAATGTTTATTCATCTCTTAGATTTCAAAAGC 3636
Qy 356 aProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHi 376
Db 3637 TCCAAATGGATCACTTGAAGCAAGAGCTCAGAAGAAATTCCTTGACGAAAAGATCA 3696
Qy 376 sArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAs 396
Db 3697 TAGCAACAAATCGATCAGACATTTACAGACATTTCTCGCGCTTTCAGTTAAACAAACCAA 3756
Qy 396 nValLeuAsnLeuLeuThrSerThrArgThrGlyGlnProLeuValAspAspTrpAs 416
Db 3757 TGTCTTAAATCTTAACTTCCACAAGAAACAGAGCAGCTCTCTGTTAGACGATTGGGA 3816
Qy 416 pCysPheLysThrLeu ----- 421
Db 3817 TTGCTTCAAGACTCTAGTAACAAACACATCTCAAACTTGTACTTGTGTTCTACGGCA 3876
Qy 422 -----ValAsnSerPh 425
Db 3877 CAACCATTTGCATTATTACTAAACCATGTATATCGAATGAAATCGCAGGTTTAAATAGCTT 3936
Qy 425 eLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAs 445
Db 3937 CAAGAATCACTGCGGTGCACAGGTCATTCAGGATTCAAGTATACAGAGCGCTTGCCAA 3996
Qy 445 nIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSe 465
Db 3997 TATCTGCAATATGGAGTGGATGTGAAGCAAACTGTTTTCAGCCATTGAACAGCTTGTTC 4056
Qy 465 rMet 466
Db 4057 GATG 4060

RESULT 3

US-09-938-842A-1485
: Sequence 1485, Application US/09938842A
: Patent No. US20020160378A1

GENERAL INFORMATION:

: APPLICANT: Harper, Jeff
: APPLICANT: Kreps, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: FILE REFERENCE: SRIPI300-3
: CURRENT APPLICATION NUMBER: US/09/938.842A
: PRIOR FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227,866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264,647
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/300,111
: PRIOR FILING DATE: 2001-06-22

```

; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1485
; LENGTH: 1485
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1485

Alignment Scores:
Pred. No.: 1,086-136
Score: 1246.00
Percent Similarity: 69.8%
Best Local Similarity: 58.5%
Query Match: 30.7%
DB: 9
Gaps: 4

US-09-934-066-2 (1-466) x US-09-938-842A-1485 (1-1485)

QY 36 GluSerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsn 55
DB 145 GAAAGGACGACGATCTTAACCTCGGTACTAGTGGGCTCTTCTAGTGGCGGATCTAGC 204
QY 56 GluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLys 75
DB 205 GCATATTGGATACAGCATCAGCATGATATATGCTATGCTATGCTATGCTATGCTATGCT 264
QY 76 GlyTyrLeuLysAspGluAsnIleValPheMetTyrAspAspIleAlaPheSerSer 95
DB 265 GGTGATTGAAGAGGAGATATGTGTTATTCATGATGATGATGATGATGATGATGATGATG 324
QY 96 GluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGly 115
DB 325 GAGATCAAGGCTGCAACCATATATCAACAGCCCTCATGGAAGATGCTCTCAGGA 384
QY 116 ValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeu 135
DB 385 GTTCCCAAGGATTATCTAGTGGATGATGATGATGATGATGATGATGATGATGATGATG 444
QY 136 GlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsp 155
DB 445 GGACACAAACTGCTGTTAAAGGGGAAGTGGGAAGTGGGAAGTGGGAAGTGGGAAGTGG 504
QY 156 AsnIlePheIleTyrAlaAspPheAsnGluValLeuGluLysMetHisLysArgLysLys 195
DB 505 CATATCTCATATCTATACAGTACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 564
QY 176 AspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLys 195
DB 565 CCTTACCTATATCAATGATCTCAATGATGATGATGATGATGATGATGATGATGATGATG 624
QY 196 TyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyTle 215
DB 625 TATAAAGCTGGTGTGTTATCTCGAAGCTTCGGAATCTGGAAGTATCTTTGAAGGGCTT 684
QY 216 LeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTyr 235
DB 685 CTCTCGAGGGTGTGAACATCTATGCAACATGATGATGATGATGATGATGATGATGATGATG 744
QY 236 GlyValTyrCysProGluSerTyrProProProProSerGluIleGlyThrCysLeuGly 255
DB 745 GTTACCTATGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 804
QY 256 AspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThr 275
DB 805 GACTTGTACAGTGTGCTGGATGAGATAGTGTGATGATGATGATGATGATGATGATGATGATG 864
QY 276 LeuGluGlnGlnTyrHisValValLysArgArg-----ValGlySerAspValPro 292
DB 865 CTCACACGATATGAACCTGTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 921
QY 293 GluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSer 312
DB 922 ---GGTCTCTATGTCATGCAATATGCGCATGTAGGAATATGCAAGGATATATCTCGATCTT 978

```

RESULT 4

```

US-09-938-842A-927
; Sequence 927, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FIELD OF INVENTION: SAME, AND METHODS OF USE
; FILING DATE: 2001-08-24
; PRIORITY DATE: 2001-08-24
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR FILING DATE: 2000-08-24
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 927
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-927

```

```

Alignment Scores:
Pred. No.: 1,426-126
Score: 1160.00
Percent Similarity: 63.6%
Best Local Similarity: 49.2%
Query Match: 47.2%
DB: 9
Gaps: 7

```

US-09-934-066-2 (1-466) x US-09-938-842A-927 (1-1461)

```

QY 7 HisPheGlnIleLeuValPheLeuHisAlaLeuIlePheSerAlaGluSerArg --- 25
DB 16 TATTTCAGACCGAGCTCTTCTTCTCTAGTCTCTTTTGGTTCATGCCGATCACCGGT 75

```

```

Qy 26 -----LysThrGlnLeuLeuAsnAspAsnValGluSer 37
Db 76 CGGTTCCAGGCAAGATCTTATCCGACGAGAGCAAGCCGCGTGAC----- 126
Qy 38 SerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyr 57
Db 127 CAAGACCAAGATGGTGCAGATGGCGGTTCTCGTGGTGTCTTCTGGGATAT 186
Qy 58 TyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGly 77
Db 187 GGAACACTACAGACACAGGCTGAGTGTGCACGCATATCAATATCAAGAAAAGGAGGT 246
Qy 78 LeuLysAspGluAsnIleIleValPheMetTyrAspIleAlaPheSerSerGluAsn 97
Db 247 TTAAGGAAGAAACATAGTCGTTTGTGATGATGATATCGCAAAACCCACCTTAAT 306
Qy 98 ProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValPro 117
Db 307 CCTCGTCCGGGTACTCTCAACCATCTCGCGGTGACGATGTTTACGCCGGAGTCCCT 366
Qy 118 LysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsn 137
Db 367 AAGGACTATAGTGTAGTACGTTACGGCTGCAAACTTCTACGCTGTACTCTAGCGGAC 426
Qy 138 GluSerGlyValThrGlyLysGlnGlyValLysValLysSerGlyProAsnAspAsnIle 157
Db 427 CAGAAGGCTGTTAAGGTGAAGCGGTAAAGTCATCGTACGAGCCCAACGATCAGATT 486
Qy 158 PheIleTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGlu 177
Db 487 TTCGTATATTATCGGATCATGGTGGTCCCGGAGTCTTGGGATGCGCAAAATACGCCCTCAC 546
Qy 178 ValMetAlaLysAspPheAsnGluValLeuGluLysMethHisLysArgLysLysTyrAsn 197
Db 547 ATATATCACGCTGATTTATTGAACGCTTAAGAAGAAGCATGCTCCGGACATACAAA 606
Qy 198 LysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLys 217
Db 607 GAGATGGTTATATACGTAGAACGCTGTGAAGTGGGAGTATTTTGAAGGGGATAATGCCA 666
Qy 218 LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyVal 237
Db 667 AAGGACTTGAACATTTACGTAAACACGCGTCAATCCACAAGAGATGTTATFGGAACA 726
Qy 238 TyrCysProGluSerTyrProProSerGluIleGlyThrCysLeuGlyAspThr 257
Db 727 TATTGTCTCGCATGAATCCGTCACCCCATCTCAATATATCATCTGCTTAGGGGATTTA 786
Qy 258 PheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGlu 277
Db 787 TATAGTCTTCTGTTGGATGGAAGATAGTGAGACTCACAAATTTAAGAAAGAGACCATTAAG 846
Qy 278 GlnGlnTyrHisValValLysArgArgValGly-----SerAspValProGluThrSer 295
Db 847 CAACAATACCACAGGCTGAGATGAGACATCAAACTACATACCTACTCAGTGGGCTCT 906
Qy 296 HisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTrpIleGly 315
Db 907 CATGTGATGAATACGGTACAATAGTATTAAAGTCGGAGAAGCTTTATCTTTACCAAGG 966
Qy 316 ArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerProIle----Ser 334
Db 967 TTTGATCCACCCCGGTTAATCTCCCACTAAACGAA-----TTACCGGTCAAGTCA 1017
Qy 335 AsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuArgLysIleGln 354
Db 1018 AATAGAGATGCTTAACCAACGCGGACCGGACCTTCTTCTTTCGGCATATGATATCGG 1077
Qy 355 LysAlaProMetGlySerLeuLeuSerLysGlnAlaGlnLysLysLeuLeuAspGluLys 374
Db 1078 ACATCGGAAGATGGTCAAGGAAGAGGATGACACATTGAAGGAATTAATGAGACAACA 1137
Qy 375 AsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGln 394

```

```

Db 1138 AGGCATAGGAACATTTAGATGCAAGCGTCAATTTGATAGCCACAATTTTGTGGTCCG 1197
Qy 395 Thr---AsnValLeuAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 413
Db 1198 ACATGATGATGTTCTTAACATTTGGTT-----AGAAACCCGGTTGGCTTTGGTTGAC 1248
Qy 414 AspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrVal 433
Db 1249 GATTGGCAATCTCTTAATTCAGTGTACGTCTATTTCAAGAGCATTTCTGGATCACTAAGC 1308
Qy 434 HisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspVal 453
Db 1309 CAATATGCGGTGAACATATCGGAGCGTTTGCACAACTTTGTAACACGCTGTGTCCAAA 1368
Qy 454 LysGlnThrValSerAlaIleGluGlnAlaCys 464
Db 1369 GAGCTGATGAGGAGGAGCTTCTACTCGCGCATGC 1401

RESULT 5
US-10-208-408-18
; Sequence 18, Application US/10208408
; Publication No. US20030096272A1
; GENERAL INFORMATION:
; APPLICANT: Scheybe, Xiao Min
; TITLE OF INVENTION: GENES REGULATED BY PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR
; FILE REFERENCE: PA-0048-1 US
; CURRENT APPLICATION NUMBER: US/10/208,408
; PRIOR FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 60/308,868
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PERL Program
; SEQ ID NO 18
; LENGTH: 1974
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030096272A1 1301477CB1
US-10-208-408-18

Alignment Scores:
Pred. No.: 6.84e-87 Length: 1974
Score: 825.00 Matches: 179
Percent Similarity: 55.13% Conservative: 79
Best Local Similarity: 38.25% Mismatches: 158
Query Match: 33.62% Indels: 52
DB: Gaps: 13

US-09-934-066-2 (1-466) x US-10-208-408-18 (1-1974)
Qy 8 PheGlnIleLeuValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr 27
Db 178 TGGAAATAGCTGTTATTCCTCAGTGTGGCCCTGGGCATTGGTGCC----- 222
Qy 28 GlnLeuLeuAsnAspAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTrp 47
Db 223 -----GTTCTATAGATGATCTCTCAAGATGGAGGCAACCACTGG 261
Qy 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
Db 262 GTGCTGATCGTGGCAGGTTCAAAATGGTGTGATATAATATAGGACCAAGGAGCGCGTGC 321
Qy 68 HisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMet 87
Db 322 CATGCCCTACCAGATCATCTACCACAATGGGATTCCTCGCAACAGATCGTTGTGATGATG 381
Qy 88 TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
Db 382 TAGCATCACATGCTTACTCTGAGACAATCCCACTCCAGGAATTTGTGATCAACAGGCC 441
Qy 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127

```

```
Db 442 AATGGCACAGATGCTATCAGGAGTCCCGAAGACTACAGTGGAGAGATGTTACCCCA 501
Qy 128 GlnAspPheTyrAsnValLeuGluValAsnGluSerGlyValThrGly---GlyAsnGly 146
Db 502 CAAATTTCTTGCTGTGTGAGAGGCGATCCAGACAGTGAAGGCGATAGATCCGC 561
Qy 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla 166
Db 562 AAGTCTGAAGAGTGGCCCGCAGATCAGCTGTTCATTTACTTCACTGACCAATGATCT 621
Qy 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
Db 622 ACTGGAAATACTGGTTTTC---AATGAAGATCTCATGTAAAGGACCTGAAATGAGACC 678
Qy 187 LeuGluLysMetHisLysArgLysTyrAsnLysMetValIleTyrValGluAlaCys 206
Db 679 ATCCATTACATGTACAAACAAATGTACCGAAGATGGTGTCTACATTGAAGCTGT 738
Qy 207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
Db 739 GAGTCTGGCTCCATGATGAACAC---CTGCCGGATAACATCAATGTTATGCAACTACT 795
Qy 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProPro 246
Db 796 GCTGCCAACCCAGAGAGTCTCTACGCTGTTACTATGATGAG-----840
Qy 247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
Db 841 -----AAGAGTCCACAGTCTGGGGAGTGGTACAGCTCACTGGATGGAGACTCG 894
Qy 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArgArg 286
Db 895 GACGTGGAAGATCTGACATAAGAGCCCTGCACAAAGAGTACACCTGGTAAATCCGAC 954
Qy 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db 955 ACC-----AACACGAGCCAGCTCATGAGTATGGAACAAACAACTCC 999
Qy 307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
Db 1000 ACCATGAAGATGATGACGTTTCAGGGT-----ATGAAA 1032
Qy 327 GluSerPheSerSer-----ProIleSerAsnSerGlyLeuValAsnPro 341
Db 1033 GCAAAAGCCAGTCTCCCGTCCCTACCTCCAGTCCACACACCTTGACCTCACCCCCAGC 1092
Qy 342 ArgAspIleProLeuTyrLeuGluArgLysIleGlnLysAlaProMetGlySerLeu 361
Db 1093 CCTGATGCTCTCACCATCATGAAAGAACTGATGACAC-----AATGATCTG 1146
Qy 362 ---GluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1147 GAGGAGTCCAGGACGCTCACGGAGGATCCAGCGGCATCTGGATGCGGACCTCAT 1206
Qy 381 AspGlnSerIleThrAspIleArgLeuGluSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1207 GAGAAATCAGTCGTAAGATCTCTCTTGTGTCGACGCGTCCGAGGCTGAGGTGGAGCAG 1266
Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db 1267 CTCCTGTCCGAGAGAGCC-----CCGCTCACGGGGCAC---AGCTCTCACCCAGAG 1314
Qy 421 LeuValAsnSerPheLysAsnHisCys-----GlyAlaThrValHisTyrGly 436
Db 1315 GCCCTGCTGCTACTCCGACCCACTGCTTCACTGGCAGCTCCCGACGTAGAGTATGCG 1374
Qy 437 LeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThr 456
Db 1375 TTGAGACATTTCTAGCTCTGCTCAACCTTTGTGAGAAGCCGTATCCGCTTCACAGATA 1434
Qy 457 ValSerAlaIleGluGlnAlaCys 464
Db 457 ValSerAlaIleGluGlnAlaCys 464
```

```
Db 1435 AATTGTCCATGGACCAAGCTGTGC 1458
RESULT 6
US-10-002-600-10
; Sequence 10 Application US/10002600
; Patent No. US20020137077A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 10
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Template ID: 1400599.2
US-10-002-600-10
Alignment Scores: 4.67e-86 Length: 2002
Pred. No.: 818.00 Matches: 179
Score: 55.22% Conservative: 80
Percent Similarity: 38.17% Mismatches: 157
Best Local Similarity: 33.33% Indels: 53
Query Match: 12 Gaps: 13
DB:
US-09-934-066-2 (1-466) x US-10-002-600-10 (1-2002)
Qy 8 PheGlnIleLeuValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr 27
Db 187 TGGAAAGTAGCTCTATCTCTCAGTGTGCCCTGGCGATTGGTGCC-----231
Qy 28 GlnLeuLeuAsnAspAsnValGluSerSerAspLysSerAlaLysGlyThrArgTrp 47
Db 232 -----GTTCTATAGATGATCCTGAAGATGGAGGCAACCACTGG 270
Qy 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
Db 271 GTGTGTATCTGTCAGATTCAAATGCTGTATATATATAGCACCGACGACCGCTGC 330
Qy 68 HisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleLeuValPheMet 87
Db 331 CATGCCCTACCAGATCATTCACCGCAATGGGATCTCCAGCAACAGATCGTTGTGATGATG 390
Qy 88 TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
Db 391 TAGCATCACATTCCTTACTCTCAAGACAATCCCACTCCAGGAATTTGTGATCAACAGGCC 450
Qy 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
Db 451 AATGGCACAGATCTCTATCAGGAGTCCCGAAGGACTACACTGGAGAGGATGTTACCCCA 510
Qy 128 GluAsnPheTyrAsnValLeuLeuGluGlyAsnGluSerGlyValThrGly---GlyAsnGly 146
Db 511 CAAATTTCTCTCTGTTGAGCGGATGAGAGGAGTGAAGGCGATAGGATCCGCC 570
Qy 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla 166
Db 571 AAGTCTGAAGAGTGGCCCGCAGATCAGCTGTTCATTTACTTCACTGACCAATGATCT 630
Qy 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
Db 631 ACTGGAATACTGTTTTC-----AATGAAGATCTTCATGTAAGAGGACCTCAATGAGACC 687
```

```
Qy 187 LeuGluLysMethHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
Db 688 ATCCATTACATGTACAAACACAAATGTACCGAAGATGCTGTCTACATTGAAGCCTGT 747
Qy 207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
Db 748 GAGCTGGGGTCCATGATGAACACAC---CTCCGGATAAACATCAATGTTTGAACACTACT 804
Qy 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProPro 246
Db 805 CTGCGCAACCCAGAGAGTCTCTAGCCCTGTACTATGATGAG-----849
Qy 247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
Db 850 -----AAGAGGTCACCGTACCTGGGGGACTGTTACAGCGTCAACTGGATGGAAGACTCG 903
Qy 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArgArg 286
Db 904 GAGCTGGAAGATCTGACTAAGAGAGCCCTGCACAAGCAGTACCACCTGGTAAATGCCAC 963
Qy 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db 964 ACC-----AACACAGCCAGCTGATGCGATGATGAACAAACAATCTCC 1008
Qy 307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
Db 1009 ACCATGAAGATGATGAGTTTCAGGT-----ATGAAA 1041
Qy 327 GluSerPheSerSerProIleSerAsnSerGlyLeuValAsnPro-----341
Db 1042 CGCAAGCAGCTGTTCCCGTCCCTACCTCCAGTCACACACACCTGACCTCACCCCCAGC 1101
Qy 342 ArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu 361
Db 1102 CTTGATGTGCTCTCACCATCATGAAGAAGAACTGATGACACC-----AATGATCTG 1155
Qy 362 ---GluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1156 GAGGAGTCCAGGAGCTCAGGAGGAGATCCAGCGGCATCTGGATGCCAGGCACCTCAT 1215
Qy 381 AspGln-SerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLe 400
Db 1216 GAGGAAGTACGTGCGTAAGATCTCTCTGCTGCGCAGCTCCGAGGCTGAGGTGAGCA 1275
Qy 400 uLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysTh 420
Db 1276 GCTCCTGTCGAGAGAGCC-----CGCTCACGGGGCAC---AGCTGCTACCCAGA 1323
Qy 420 rLeuValAsnSerPheLysAsnHisCys-----GlyAlaThrValHisTyrG1 436
Db 1324 GGCCCTGCTGCACTCCGGACCCACTGCTCAACTGGCACTCCCCACGCTACGAGTATGC 1383
Qy 436 yLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnTh 456
Db 1384 GTTGAGACATTTGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1443
Qy 456 rValSerAlaIleGluGlnAlaCys 464
Db 1444 AANAATTCCATGGACCACCTGTGC 1468
```

RESULT 7

```
US-09-967-796-1
; Sequence 1, Application US/09967796
; Patent No. US2002015535A1
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; Braxton, Scott M.
; Deleane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
```

```
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/967,796
FILING DATE: 28-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/449,422
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33954
REFERENCE/DOCKET NUMBER: PF-0048 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1855 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: Adrenal
CLONE: 100877
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-967-796-1
```

```
Alignment Scores:
Pred. No.: 2,17e-77 Length: 1855
Score: 744.00 Matches: 166
Percent Similarity: 54.63% Conservative: 70
Best Local Similarity: 38.43% Mismatches: 128
Query Match: 30.32% Indels: 68
DB: Gaps: 14
```

```
US-09-934-066-2 (1-466) x US-09-967-796-1 (1-1855)
Qy 8 PheGlnIleLeuValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr 27
Db 60 TGGAAAGTAGTTGTTATCTCTCAGTGTGGCCCTGGGAATTGGTGCC-----104
Qy 28 GlnLeuLeuAsnAspAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTyr 47
Db 105 -----GTTCTATAGATGATCTCTGAAGTGGAGCAGCAGCTGG 143
Qy 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
Db 144 GTGGTATCGTGGCAGGTTCAATGCTGCTATATATAGGCACCCAGCAGCAGCGCTGC 203
Qy 68 HisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMet 87
Db 204 CATGCCCTACCAGTTTCATTCCACCCAATGGGATTCCTGCCGACAGATCGTTGTGATTATG 263
Qy 88 TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
Db 264 TAGCATGACATAGCTTACTCTGAGACACATCCCACTCCAGCAATGTGATCAACAGCCCC 323
Qy 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
Db 324 AATGGCAGACATGCTATCAGGAGTCCCGCAGGAGTACACTGAGAGGATGTTATCCCA 383
Qy 128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly----GlyAsnGly 146
Db 128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly----GlyAsnGly 146
```

```

Db 384 CAAAATTTCCTGTTGTTGAGAGCGGATGAGAGCGGATGAGAGGTTATAGGATCCCGC 443
Qy 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrAlaAspHisGlyAla 166
Db 444 AAAGTCTGAAGAGTGGTCCCGAGATCAGCTGTTTATTTTCTACTGACCATGGATCT 503
Qy 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
Db 504 TCTGGAATAGTCTGTTTCCCTCC---AATGAAGATCTTCTGTAAGAGGACCTGATTAGACC 560
Qy 187 LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
Db 561 ACCAFTTACATTTTCAAAACAAATGATCGAAGAGTGTCTTACTATGAGGCTGT 620
Qy 207 GluSerGlySerMetPheGluGlyLysLysAsnLysLysValIleTyrAlaValThr 226
Db 621 GAGTCTGGTCTCATGATGAACAC---CTGCCGATACATCAATGTTTATGCACTACT 677
Qy 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProPro 246
Db 678 GCTGCCAACCCAGAGATCTCTCTAGCCTGTACTATGATGAG--- 722
Qy 247 ProSerGluIleGlyThrCysLysLysLysLysLysLysLysLysLysLysLysLys 266
Db 723 -----AAGAGTCCACGTACCTGGGGACTGGTACAGCTCAACTGGATGGAACATCG 776
Qy 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArg 286
Db 777 GAGTGGGAAGATCTGACTAAGAGACCTGCACAAGCAGTACCACTGTTGTAATAATCGCAC 836
Qy 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db 837 ACC-----AACACCCAGCAGTCACTGATGGAACAAACAAATCTCC 881
Qy 307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPhe 326
Db 882 ACCATGAAGATGATGCGAGTTTACGGT-----ATGAAA 914
Qy 327 GluSerPheSer-----ProIleSerAsnSerGlyLeuValAsnPro 341
Db 915 CGCAAGCCAGTCTCCCGTCCCTACCTCCAGTCAACACCTTGACCTCACCCCGCAGC 974
Qy 342 ArgAspIleProLeuGlnArgLysLysLysLysLysLysLysLysLysLysLysLys 361
Db 975 CCTGATGCTCTCACCATCATGAAGAAAGCAACTGATGACAC---AATGATCTG 1028
Qy 362 GluSerLys-----GluAlaGlnLysLysLysLysLysLysLysLysLysLysLys 377
Db 1029 GAGGAGTCCAGGAGCTCAGGAGGAGATCCAGCGGTATCTGGATGCCAGCAC----- 1082
Qy 378 LysGlnIleAspGlnSerIleThrAspIleThrAspIleThrAspIleThrAspVal 397
Db 1083 -----CTCATCCGAGGTGAGGTGAGGACAG----- 1106
Qy 398 LeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCys 417
Db 1107 -----CTCCTGTCCAGAGACCC-----CCGCTCAGGCGGCAC---ACCTCC 1145
Qy 418 PheLysThrLeuValAsnSerPheLysAsnHisCys 429
Db 1146 TACCCAGAGGTCTCTGTTACTTCCGACCCACTGC 1181

```

RESULT 8

```

US-10-043-487-97
Sequence 97, Application US/10043487
Publication No. US20030055220A1
GENERAL INFORMATION:
APPLICANT: HIRAGENICS
APPLICANT: HIRAGENICS
TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptides
TITLE OF INVENTION: mammalian polypeptides
FILE REFERENCE: PA 778A
CURRENT APPLICATION NUMBER: US/10/043.487

```

```

: CURRENT FILING DATE: 2002-04-30
: PRIOR APPLICATION NUMBER: US 60/261,130
: PRIOR FILING DATE: 2001-01-12
: NUMBER OF SEQ ID NOS: 561
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 97
: LENGTH: 960
: TYPE: DNA
: ORGANISM: Shigella Flexneri
US-10-043-487-97

```

```

Alignment Scores:
Pred. No.: 1,46e-47 Length: 960
Score: 487.00 Matches: 115
Percent Similarity: 52.19% Conservative: 64
Best Local Similarity: 33.53% Mismatches: 124
Query Match: 19.85% Indels: 40
DB: 9 Gaps: 12

```

US-09-934-066-2 (1-466) x US-10-043-487-97 (1-960)

```

Qy 133 ValLeuLeuGlyAsnGluSerGlyValThrGly---GlyAsnGlyLysValValLysSer 151
Db 4 GTGTTGAGAGGGGATGCGAAGCAGTGAAGGCGCATAGGATCCGCAAGTCTCTGAAGACT 63
Qy 152 GlyProAsnAspAsnIlePheIleTyrAlaAspHisGlyAlaProGlyLeuIleAla 171
Db 64 GCGCCCGCAGGATCAGCTGTTTACTTCTACTGACCATGATCTACTGCAATACTGGTT 123
Qy 172 MetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHis 191
Db 124 TTTTCCC---AATGAAGATCTTCTGTAAGGACCTGAATGAGACCATCCATTATCATGTAC 180
Qy 192 LysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMet 211
Db 181 AAACCAAAATGTACCGAAGATGTTCTTCTACTGAAAGCCTGTGAGTCTGGGTCCATG 240
Qy 212 PheGluGlyIleLeuLysLysAsnLysLysLysLysLysLysLysLysLysLysLys 231
Db 241 ATGAACACAC---CTGCCGGATACATCAATGTTTATGCACTACTGCTGCCAACCCAGA 297
Qy 232 GluSerSerTrpGlyValTyrCysProGluSerTyrProProProProSerGluIleGly 251
Db 298 GAGTCTGCTTACGCTTCTTACTATGATGAG-----AAGAGGTCC 336
Qy 252 ThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMet 271
Db 337 ACGTACCTGGGGACTGGTACACGCTCACTGGTGGAGACTCGGACGTGGAAGATCTG 396
Qy 272 SerLysGluThrLeuGluGlnTyrHisValValLysArgValGlySerAspVal 291
Db 397 ACTAAGAGACCTTGCACAAGCAGTACCACCTGGTAAATCCACAC--- 444
Qy 292 ProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSer 311
Db 445 ---AACACACCGCCGCTGATGCAATGCAAAACAAACAACTCCACCCTGAAAGTATG 501
Qy 312 SerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSer 331
Db 502 CAGTTTCAGGGT-----ATGAACGCAAGACCGCTTCT 534
Qy 332 -----ProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeu 346
Db 535 CCGCTCCCGCTACCTCCAGTCAACACCTTGACCTCCACCCCGCCTGTGCTCTC 594
Qy 347 LeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu---GluSerLysGlu 365
Db 595 ACCATCATGAAAGGAAACTGATGAAACAC-----AATGATCTGGAGGAGTCCAGGCAG 648
Qy 366 AlaGlnLysLysLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThr 385
Db 649 CTCACGAGGAGATCCAGCGCATCTGATGCCAGGCACCTCATTTGAGAAGTCACTGCGT 708

```


TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
 FILE REFERENCE: 16511.006/37-21(10298)C

US-09-934-066-2 (1-466) x US-09-960-352-11028 (1-422)

```
Qy 29 LeuLeuAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTirpAla 48
Db ::::|
94 GTCTGGGAACGTGTGCTGCCCTCGAGGACCTCGAGGACGAGGCAAGCACTGGTG 153
Qy 49 ValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHis 68
Db ::::|
154 GTGATCGTCGAGGATCAACAGCGTGTATTAATACAGGACCAAGGAGGATGATGCCAT 213
Qy 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
Db ::::|
214 GCCTACCATGCTGTACCAAGGATCCCTGATGAGCAGATCATCGTGATGATGAT 273
Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db ::::|
274 GAGCAGATCGCAACTCCGAAGACAAATCCACCCCGGGAATGTGATCAACAGGCCCAAC 333
Qy 109 GlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnValGln 128
Db ::::|
334 GGCTCAGACGTGTACCAAGCGTGTGTAAGGACTACACGGCGGAAGATGTACCCCGAAG 393
Qy 129 AsnPheTyrAsnValLeuLeuGlyAsn 137
Db ::::|
394 AATTTCCTGCTGTGTGAGAGGTGAT 420
```

RESULT 14

US-09-960-352-1596
: Sequence 1596, Application US/09960352
: Patent NO. US20020137139A1

GENERAL INFORMATION:

: APPLICANT: Warren, Wesley C.

: APPLICANT: Tao, Mengbing

: APPLICANT: Byatt, John C.

: APPLICANT: Mathialagan, Nagappan

: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

: FILE REFERENCE: 16511.006/37-21(10298)C

: CURRENT APPLICATION NUMBER: US/09/960.352

: CURRENT FILING DATE: 2001-09-24

: NUMBER OF SEQ ID NOS: 15112

: SEQ ID NO 1596

: LENGTH: 426

: TYPE: DNA

: ORGANISM: Bos taurus

: OTHER INFORMATION: Clone ID: 07-LIB34-074-Q1-E1-B3

US-09-960-352-1596

Alignment Scores:

Pred. No.: 5.62e-30 Length: 426
Score: 333.00 Matches: 62
Percent Similarity: 69.72% Conservative: 14
Best Local Similarity: 56.88% Mismatches: 33
Query Match: 13.57% Indels: 0
DB: 10 Gaps: 0

US-09-934-066-2 (1-466) x US-09-960-352-1596 (1-426)

```
Qy 29 LeuLeuAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTirpAla 48
Db ::::|
98 GTCTGGGAACGTGTGCTGCCCTCGAGGACCTCGAGGACGAGGCAAGCACTGGTG 157
Qy 49 ValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHis 68
Db ::::|
158 GTGATCGTCGAGGATCAACAGCGTGTATTAATACAGGACCAAGGAGGATGATGCCAC 217
Qy 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
Db ::::|
218 GCCTACCATGCTGTACCAAGGATCCCTGATGAGCAGATCATCGTGATGATGAT 277
Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db ::::|
```

Search completed: May 29, 2003, 20:43:48

Job time : 209 secs

```
Db 278 GAGCAGATCGCCAACTCCGAAGACAATCCACCCCGGAATTTGTATCAACAGGCCCAAC 337
Qy 109 GlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnValGln 128
Db ::::|
338 GGCTCAGACGTGTACCAAGCGTGTCTGAAGGACTACACGGCGGAGGATGTACCCCCGAAG 397
Qy 129 AsnPheTyrAsnValLeuLeuGlyAsn 137
Db ::::|
398 AATTTCCTGCTGTGTGAGAGGTGAT 424
```

RESULT 15

US-09-960-352-5110

: Sequence 5110, Application US/09960352

: Patent NO. US20020137139A1

: GENERAL INFORMATION:

: APPLICANT: Warren, Wesley C.

: APPLICANT: Tao, Mengbing

: APPLICANT: Byatt, John C.

: APPLICANT: Mathialagan, Nagappan

: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

: FILE REFERENCE: 16511.006/37-21(10298)C

: CURRENT APPLICATION NUMBER: US/09/960.352

: CURRENT FILING DATE: 2001-09-24

: NUMBER OF SEQ ID NOS: 15112

: SEQ ID NO 5110

: LENGTH: 407

: TYPE: DNA

: ORGANISM: Bos taurus

: OTHER INFORMATION: Clone ID: 22-LIB34-005-Q1-E1-F5

US-09-960-352-5110

Alignment Scores:

Pred. No.: 6.05e-27 Length: 407
Score: 307.00 Matches: 55
Percent Similarity: 72.04% Conservative: 12
Best Local Similarity: 59.14% Mismatches: 26
Query Match: 12.51% Indels: 0
DB: 10 Gaps: 0

US-09-934-066-2 (1-466) x US-09-960-352-5110 (1-407)

```
Qy 29 LeuLeuAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTirpAla 48
Db ::::|
124 GTCTGGGAACGTGTGCTGCCCTCGAGGACCTCGAGGACGAGGCAAGCACTGGTG 183
Qy 49 ValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHis 68
Db ::::|
184 GTGATCGTCGAGGATCAACAGCGTGTATTAATACAGGACCAAGGAGGATGATGCCAT 243
Qy 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
Db ::::|
244 GCCTACCATGCTGTACCAAGCGGATTCCTGATGAGCAGATCATCGTGATGATGAT 303
Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db ::::|
304 GAGCAGATCGCCAACTCCGAAGACAATCCACCCCGGAATTTGTATCAACAGGCCCAAC 363
Qy 109 GlyGluAspValTyrLysGlyValProLysAspTyrThr 121
Db ::::|
364 GGCTCAGACGTGTACCAAGCGGCTGTGTAAGGAGCTACACG 402
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 18:21:49 ; Search time 2474 Seconds
(without alignments)
4735.812 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHPQLVFLHALLIF.....CNMGVDVKTVAIEQACSM 466

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delopt 6.0 , Delext 7.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlp

-O=/cgn2_1/USPTO.spool/US0934066/runat_19052003_163701_24421/app_query.fasta_1.647
-DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=p2n.rnnp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -XPRY=US0934066 -CGN_1_1_4447 -runat_19052003_163701_24421
-NCPU=6 -ICPU=3 -NO_XLPR -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT
-DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7 -GGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq.*

29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq.*
44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2454	100.0	1401	19	US-09-513-996A-66846
2	2454	100.0	1560	35	US-09-934-066-1
3	2454	100.0	1560	66	US-60-226-804-1
4	2446	99.7	1541	19	US-09-513-996A-11859
5	2188	89.2	4320	35	US-09-934-066-3
6	2188	89.2	4320	66	US-60-226-804-3

Db 1021 CCSCGGCATATCTCTGCTATACCTCCAGAGAAAGATTCAAAAAGCTCCAATGGGATCA 1080
Qy LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1081 CTTGAAGCAAGAGCTCAGAGAATCTCTGACGAAAGAAATCATAGCAACAATC 1140
Qy AspGluSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1141 GATCAGAGCATTCAGACATCTCGCGCTTTCAGTTAAACCAACCAATGCTTAAATCTC 1200
Qy LeuThrSerThrArgThrGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db 1201 TTAACCTTCCAGAACACAGACAGCCCTCTGTAGACATTGGGATTGCTTCAAGACT 1260
Qy LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db 1261 CTAGTTAATAGCTCAAGATCACTCGGCTGCAACGGTGATTCAGGATTGAAGTATACA 1320
Qy GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db 1321 GGAGCGCTTCCCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTCAGCCATT 1380
Qy 461 GluGlnAlaCysSerMet 466
Db 1381 GAACAAGCTTGTTCGATG 1398

RESULT 2
US-09-934-066-1
: Sequence 1. Application US/09934066
: GENERAL INFORMATION:
: APPLICANT: Gruis, Darren B.
: APPLICANT: Jung, Rudolf
: TITLE OF INVENTION: Methods of Increasing Polypeptide
: FILE REFERENCE: 35718/237251
: CURRENT APPLICATION NUMBER: US/09/934,066
: CURRENT FILING DATE: 2001-08-21
: PRIOR FILING DATE: 2001-08-21
: PRIOR FILING DATE: 2000-08-21
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 1560
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-934-066-1

Alignment Scores:
Pred. No.: 4,7e-234 Length: 1560
Score: 2454.00 Matches: 466
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 35 Gaps: 0

US-09-934-066-2 (1-466) x US-09-934-066-1 (1-1560)

Qy 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 40 ATGCTAGTCCTCTGGTCACTTCAGATTCCTGTTTCTTCATGCTTGTCTATCTTC 99
Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspValGluSerSerAspLys 40
Db 100 TCAGCTGAGTCCGCAAAACCCCAATGCTCAACGATAATGATCTGAATCTACGCAAG 159
Qy 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db 160 AGTGCAAAAGCCACGATGGCTGTTTATGCTGGATCAATGAATATTAATACTAC 219
Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp 80
Db 220 AGGCATCAGGCTGACATATGCCACGGGTATCAGATACTCCGAAAAGCGGCTTTAAAGAT 279

Qy 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 280 GAAACATCATCTGTTTATGATGATATCGCGTTTTTCTCGGAGAATCTTAGSCCT 339
Qy 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 340 GGAGTTTATCATTAATAAACACAGATGGAGAAGATGTTTATAAAGGAGTCTTAAGGACTAC 399
Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 400 ACTAAGAAGCTGTTAATGTTCAAACTCTACAACTGTGTACTTGGAAATGAAGTGGC 459
Qy 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db 460 GTCACAGGAGAAATGCAAAAGTTGTGAAAGTGGTCTAATGATAATATCTTCATCTAT 519
Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db 520 TATGCTGACCATGGAGCTCCTGCTTAATAGCCATGCCACTGGTGATGAAGTTATGGCA 579
Qy 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200
Db 580 AAAGATTCAATGAAGCTTTGGAGAAGATGCATAAGAGAAAAAATACACAAGATGGTG 639
Qy 201 IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db 640 ATCTATGTTGAAGCATGTGAATCAGCAAGTATCTTTGAAGGGATTTTAAAGAAAAATCTC 699
Qy 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPro 240
Db 700 ACATATACCGACGTGACTGCTGCTAATCTTAAAGAGAGCAGCTGGGAGTTTACTGCTCT 759
Qy 241 GluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db 760 GAGTCATATCTCTCTCTCTCTGAGATGGAACTGTCTCGGCGCATACATTTAGCATC 819
Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
Db 820 TCTTGGCTTGAGGACAGTGCCTTCATGACATGAGCAAGAGAGACTTTGGAGCAACAATAC 879
Qy 281 HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db 880 CAGTTGTAAAGAGAGAGTAGATCTGATGTACCAGAGACTTCTCATGTATCCGCTTTC 939
Qy 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
Db 940 GGAACAGAGAGATGCTTAAAGATTATCTTCTCTTACATTGGAAGAAATCTCGAAAC 999
Qy 321 AspAsnPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsn 340
Db 1000 GATAACTTCACTTTCACGGAATCCTTTTCTCACCATCTCTAATCTGGCTTTGGTCAAT 1059
Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db 1060 CCGCGCGATATCTCTGCTATACCTCCAGAGAAAGATTCAAAAAGCTCCAATGGGATCA 1119
Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1120 CTTGAAGCAAGAGAGCTCAGAGAATTCCTGACGAAAGAAATCATAGCAACAATC 1179
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1180 GATCAGAGCATTCAGACATCTCGCGCTTTCAGTTAAACCAACCAATGCTTAAATCTC 1239
Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db 1240 TTAACCTTCCAGAACACAGACAGCCCTCTGTTGACAGCATTTGGGATGCTTCAAGACT 1299
Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db 1300 CTAGTTAATAGCTTCAAGAAATCACTCGGCTGCAACGGTGATTCAGGATTAAGATACA 1359
Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460

```
Db 1360 GGAGCCCTGGCAATATCTGCAATATGGAGTGGATGTGAAGCAAACTGTTTCAGCCATT 1419
Qy 461 GluGlnAlaCysSerMet 466
Db 1420 GAACAGCTTGTCGATG 1437

RESULT 3
US-60-226-804-1
; Sequence 1, Application US/60226804
; GENERAL INFORMATION:
; APPLICANT: Jung, Rudolf
; TITLE OF INVENTION: Methods of Increasing Polypeptide
; FILE OF INVENTION: Accumulation in Plants
; FILE REFERENCE: 5718-134P
; CURRENT APPLICATION NUMBER: US/60/226,804
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-60-226-804-1

Alignment Scores:
Pred. No.: 4,7e-234 Length: 1560
Score: 2454.00 Matches: 466
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 66

US-09-934-066-2 (1-466) x US-60-226-804-1 (1-1560)
Qy 1 MetSerSerProLeuGlyHisPheGlnIleValPheLeuHisAlaLeuLeuIlePhe 20
Db 40 ATGCTAGTCCCTGGTGCACATTCAGATTCCTGTTTCTTCATGCTTTCGCTTATCTTC 99
Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAspLys 40
Db 100 TCAGCTGAGTCCGCAAAACCCCAATGCTGCAACGATAATGATGTAATCAGCAGCAAG 159
Qy 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyTyAsnTyr 60
Db 160 AGTGCAAAAGGCACACGATGCGCTGTTTGTAGTGTGCTGATCAATCAATATATAACTAC 219
Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrglnIleLeuArgLysGlyLeuLysAsp 80
Db 220 AGGCATCAGCTGACATATGCCACGCTATCAGATACTCCGAAAGCGGTTTAAAGAT 279
Qy 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 280 GAAACATCATTTGTTTATGTATCATCATATCGCGTTCCTCGAGAAATCCTAGCGCT 339
Qy 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 340 GGAGTATCATTAATAACACAGATCGAGACAGATGTTTATAAGAGCTTCCTAAGCATAC 399
Qy 121 ThrLysGluAlaValAsnValGlnAsnPhetyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 400 ACTAAAGAGCTGTTAATGTTCAAACTCTACATGTTTACTGTGAATGAAGTGGC 459
Qy 141 ValThrGlyCysAlaCysLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db 460 GTCACAGGAGGAATGGCAAGTCTGAAAGTGGTCTCAATGATAATATCTTCATCAT 519
Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db 520 TATGCTGACCATGGAGCTCTGCTGCTTAATAGCATGCCACTGTGTGATGATGATGCA 579
Qy 181 LysAspPheAsnGluValLeuGluLysMethIleLysArgLysLysTyrAsnLysMetVal 200
```

```
Db 580 AAAGATTTCAATGAAGTCTTGGAGACATCCATAAGAGAAAAAATACACAAGATGGTC 639
Qy 201 IleTyrValGluAlaCysGluSerClySerMetPheGlyIleLeuLysLysAsnLeu 220
Db 640 ATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTTCAGGGATTTTAAAGAAAAATCTC 699
Qy 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPro 240
Db 700 AACATATACGAGTGCAGTGCCTGCTAAATCTAAAGAGACAGCTGGGGAGTTTACTCTCT 759
Qy 241 GluSerTyrProProProProSerGluIleClyThrCysLeuGlyAspThrPheSerIle 260
Db 760 GAGTCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 819
Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
Db 820 TCTTGGCTTGAGGACAGTGCACCTTTCATCACAATGAGCAAGAGACTTTGGAGCAACAATAC 879
Qy 281 HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db 880 CAGCTTGTAAGAGAGAGAGTAGGATCTGATGTACCAGAGACTTCTCATGTATCCCGTTTC 939
Qy 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
Db 940 GGAACAGAGAGAGAGTAAAGATTATCTTCTCTTACATTCGAAGAAATCTCGAAAC 999
Qy 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db 1000 GATAACTTCACTTTCACGGAATCTCTTCTCTCAACAATCTCTAATCTGGCTTGGTCAAT 1059
Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db 1060 CCGCGCATATCTCTCTCTATATCTCCAGAGAAAGATTCAAAAAGCTCCAAATGGGATCA 1119
Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1120 CTTGAAACCAAGAGAGCTCAGAGAAATTCCTTGACGAAAGAAATCATAGGAAACAAATC 1179
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1180 GATCAGAGCATTTACAGACATTCCTCGCGCTTTCAGTTTAAACAAACCAATGTCTTAAATCTC 1239
Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db 1240 TTAACCTCCACAAGAACCAACAGAGAGAGCTCTTGTAGACGATTGGGATTGCTTCAAGACT 1299
Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db 1300 CTAGTTAATAGCTTCAAGAAATCACTGCGGTGCAACGCTGCATTACGCAATGGAAGTATACA 1359
Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db 1360 CGAGCGCTTGCCATATCTGCAATATGGAGTGGATCTGAAGCAAACTGTTTCAGCCATT 1419
Qy 461 GluGlnAlaCysSerMet 466
Db 1420 GAACAGCTTGTTGATG 1437

RESULT 4
US-09-513-996A-11859
; Sequence 11859, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513.996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 11859
; LENGTH: 1541
; TYPE: DNA
```

```

: ORGANISM: Arabidopsis thaliana
: FEATURE:
: NAME/KEY: UNSURE
: LOCATION: 1..1541
: OTHER INFORMATION: any n or xaa = unknown
: FEATURE:
: OTHER INFORMATION: Location 1..1541 / Ceres Seq. ID 1375309
US-09-513-996A-11859

Alignment Scores:
Pred. No.:      2,91e-233      Length:      1541
Score:          2446.00        Matches:      465
Percent Similarity: 99.79%      Conservative: 0
Best Local Similarity: 99.79%      Mismatches: 1
Query Match:      99.67%      Indels:      0
DB:              19          Gaps:        0

US-09-934-066-2 (1-466) x US-09-513-996A-11859 (1-1541)

Qy 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 59 ATGCTAGTCTCTTGCTCACTTTCAGATTCTTGTCTTCTTCATGCTTTGCTTATCTTC 118
Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAspValGluSerSerAspLys 40
Db 119 TCAGCTGAGTCCGCAAAACCCAAATTCGTAACGATAATGATGTTGAATCTAGCGACAAG 178
Qy 41 SerAlaLysGlyThrArgTTPAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db 179 AGTCAAAAGGCACACATGGCTGTTTGTAGTGTGATCAAAATGAATATTAAGACTAC 238
Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp 80
Db 239 AGGCATCAGGCTGACATATGCCACGCGTATCAGATATCTCCGAAAGCGGCTTTAAAGAT 298
Qy 81 GluAsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 299 GAAACATCATGTGTTTATGTATGATGATATCGCTTTTCCCTCGGAGAAATCTTAGGCCT 358
Qy 101 GlyValIleLeuAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 359 GGAGTTATCATTAATAACACAGATGGAGAGATGTTTATAAGGAGTTCCTAAGGACTAC 418
Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 419 ACTAAGAAGCTGTTAATGTTTCAAACTTCTACAATGTGTTACTTGGAAATGAAAGTGC 478
Qy 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db 479 GTACAGGAGAAATGGCAAGTTGTGAAAGTGGTCTTAATGATAATATCTTCATCTAT 538
Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db 539 TATGCTGACCATGGAGCTCTGCTGCTTAATAGCGATGCCACTGCTGATGAAGTATGGCA 598
Qy 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200
Db 599 AAAGATTTCATGAAGTCTTGGAGAACATGATAAGAGAAAAATATACAAACAGATGCTG 658
Qy 201 IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db 659 ATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAAATCTC 718
Qy 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerTrpGlyValTyrCysPro 240
Db 719 AACATATACCGATGACTGCTGCTAATCTTAAGAGAGACAGCTGGGAGTTACTGTCCT 778
Qy 241 GluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db 779 GAGTCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 838
Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
Db 838 TCTTGGCTTGAGGACAGTGCACCTTCATGACATGAGCAAGAGACTTTTGGACAAACAATAC 898
Qy 281 HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db 899 CACCTTGTAAAGAGAGAGTAGGATCTGATGTACCAAGAGACTTCTCATGTATGCGCTTC 958
Qy 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
Db 959 GGAACAGAGAGATGCTTAAAGATTTATCTTCTCTTACATTTGCAAGAAATCTCTGAAAC 1018
Qy 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db 1019 GATAAATCTCACTTTCACGGAATCTTTTCTCTCAACATCTCTAATCTGCGTTGGTCAAT 1078
Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db 1079 CCGCGCGATATTCCTCTGCTATACCTCCAGAGAAGATTCCAAAAGCTCCAAATGGGATCA 1138
Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1139 CTTGAAAGCAAAAGAGCTCAGAAAGATGCTTGACCAAAAGAAATCATAGCAAAACAATC 1198
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1199 GATCAGAGCATTCACAGACATTCGCGCTTTCAGTTAAACAAACCAATGTCTTAAATCTC 1258
Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTyrAspCysPheLysThr 420
Db 1259 TTAACITCCAGAGAACACAGACAGACGCTCTTGTAGACGATGGGATTCCTCAAGACT 1318
Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db 1319 CTAGTTAATAGCTTCAAGAAATCACTGCGGTGCAACGCTGCATTACGGATTGAAGTATACA 1378
Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db 1379 GGAGCGCTTCCCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTTCAGCCAT 1438
Qy 461 GluGlnAlaCysSerMet 466
Db 1439 GAACAAGCTTGTCTGATG 1456

RESULT 5
US-09-934-066-3
: Sequence 3, Application US/09934066
: GENERAL INFORMATION:
: APPLICANT: Gruis, Darren B.
: TITLE OF INVENTION: Methods of Increasing Polypeptide
: FILE REFERENCE: 35718/237251
: CURRENT APPLICATION NUMBER: US/09/934,066
: CURRENT FILING DATE: 2001-08-21
: PRIOR APPLICATION NUMBER: US 60/226,804
: PRIOR FILING DATE: 2000-08-21
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 4320
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-934-066-3

Alignment Scores:
Pred. No.:      6.43e-207      Length:      4320
Score:          2188.00        Matches:      463
Percent Similarity: 67.89%      Conservative: 0
Best Local Similarity: 67.89%      Mismatches: 3
Query Match:      89.16%      Indels:      219
DB:              35          Gaps:        7

US-09-934-066-2 (1-466) x US-09-934-066-3 (1-4320)
```



```

QY 422 -----ValAsnSerPh 425
Db 3877 CAACATTGCATTATTACTAAACAGTGTATATCGAATGAAATCGCAGGTTAATAGCTT 3936
QY 425 eLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAs 445
Db 3937 CAAGAATCACTCGGTGCAACGGTGCATTTACGGATTGAAGTATACAGGCGCTTGCCAA 3996
QY 445 nIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGlnAlaCysSe 465
Db 3997 TATCTGCAATATGGAGTGGATGTGAACCAACACTGTTTCAGCCCAATGAACAGCTGTTC 4056
QY 465 rMet 466
Db 4057 GATG 4060

RESULT 7
US-09-534-859-443
: Sequence 443, Application US/09534859
: GENERAL INFORMATION:
: APPLICANT: Bush, David F.
: APPLICANT: Last, Robert L.
: APPLICANT: Levin, Irena M.
: APPLICANT: Morris, Susan R.
: APPLICANT: Parnell, Laurence D.
: APPLICANT: Rounsley, Steven D.
: APPLICANT: Wiegand, Roger C.
: TITLE OF INVENTION: PLANT POLYMORPHIC MARKERS AND USES THEREOF
: FILE REFERENCE: 38-10(15493)B
: CURRENT APPLICATION NUMBER: US/09/534,859
: CURRENT FILING DATE: 2000-03-29
: NUMBER OF SEQ ID NOS: 1127
: SEQ ID NO 443
: LENGTH: 83253
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-534-859-443

Alignment Scores:
Pred. No.: 3,266-205 Length: 83253
Score: 2188.00 Matches: 463
Percent Similarity: 67.89% Conservative: 0
Best Local Similarity: 67.89% Mismatches: 3
Query Match: 89.16% Indels: 219
DB: 20 Gaps: 7

US-09-934-066-2 (1-466) x US-09-534-859-443 (1-83253)
QY 1 MetSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 72700 ATGTCTAGTCCCTTGGTCACTTTCAGATCTCTGTTTCTTCATGCTTTGCTTATCTTC 72759
QY 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAspValGluSerSerAspLys 40
Db 72760 TCAGCTGAGTCCCGCAACCCCAATGTCTGAACGATAATGATGTTGAATCTAGCGACAAG 72819
QY 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTrpAsnTyr 60
Db 72820 AGTGCAAAAGGCACACGATGGCTGTTTAGTCTCGTGATCAATGAATATTATAACTAC 72879
QY 61 ArgHis----- 62
Db 72880 AGGCATCAGGTTGTTAAATTATGTTGAACGTTTAAACATAACAAAAAAGGTCCA 72939
QY 63 -----GlnAlaAspIleCysHis 68
Db 72940 AGCGAGATTGTTGTAACATAAATCGACCGAGCTTTTATTTTCACAGGCTGACATATGCCAC 72999
QY 69 AlaTrpGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
Db 73000 GCGTATCAGATACTCCGAAGGCGGTTTAAAGATCAACATCAATGTTGTTTATGATAT 73059

```

```

QY 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db 73060 GATGATATCGGCTTTCTCGGAGAATCCTAGGCTGGAGTTATCATTAATAAACAGAT 73119
QY 109 GlyGluAspValTyrLysGlyValProLys----- 118
Db 73120 CGAGAAGATGTTTATAAAGAGTTCCCTAA--GGTCTTATTCTTACTCTTTCTGCGCTTA 73178
QY 118 ----- 118
Db 73179 TTTCTACCTTCAATCAATTCATATATATATATCAAGTTTGTGTTTGTGTTGGGTAG 73238
QY 119 AspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGlu 138
Db 73239 GACTACACTAAAGAAGCTGTTAATGTTCAAAACTTCTACAAATGTGTACTTGGAAATGAA 73298
QY 139 SerGlyValThrGlyLysValValLysSerGlyProAsnAspAsnIlePhe 158
Db 73299 AGTGGCGTCAACAGGAGGAAATGCGCAAGTTCTGAAAAGTGGCTCTAATGATATATCTTC 73358
QY 159 IleTyrTyrAlaAspHisGlyAlaProGlyLeuIle----- 170
Db 73359 ATCTATTATGCTGACCATGGAGCTCCTGGCTTAATAGGTTTTCTTAATTTTATGAAATTA 73418
QY 170 ----- 170
Db 73419 TTACGTACCATCAATCCATATATATATAAAGATTTTCTTTGATACATACGAAACCGG 73478
QY 171 -----AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeu 187
Db 73479 ATTTTCTCAGCGATGCCACATGGTGTGAAGTATATGCAAAAGATTTCAATGAAGTCTTG 73538
QY 188 GluLysMetHisLysArgLysLysTyrAsnLysMetVal----- 200
Db 73539 GAGAAGATGCATAAGAGAAAAAATACAAAGATGGT--ATATAACTCAACCATTCCTTTA 73597
QY 200 ----- 200
Db 73598 CCTAGCTTTATACATATGCTGTTCTGTTTGTGAATCTCTATGTTGTTGTTTGTGATGTT 73657
QY 201 -----IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLys 218
Db 73658 TAGGTGATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTCAAGGATTTTAAAGAAA 73717
QY 219 AsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerTrpGlyValTyr 238
Db 73718 AATCTCAACATATACGCAGTGACTGCTGCTAATTTCTAAACAGAGCAGCTGGGCGATTAC 73777
QY 239 CysProGluSerTyrProProProSerClnIleGlyThrCysLeuGlyAspThrPhe 258
Db 73778 TGCTCTGAGTCAATATCCTCTCCTCCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 73837
QY 259 SerIleSerTrpLeuGlu----- 264
Db 73838 AGCATCTCTTGGCTTGAGGACAGGTACTGCAAAACAAAAAGATTCAATCCTTTATGSACTA 73897
QY 264 ----- 264
Db 73898 TTGCAATGATTGATTGTTTCTTGAAGAATATTTGTTTCAATTTGTTTCTATGTTTGTGTT 73957
QY 265 -----AspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisVa 282
Db 73958 GTTTGGGACAGTCACCTTTCATGACATGAGCAAAAGACATTTTGGAGCAACAATACCAGT 74017
QY 282 lValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPheGlyThr 302
Db 74018 TGTAAAGAGAAGACTAGTAGTCTGATGTACCAGAGACTTCTCATGTATGCGCTTTCGGAAC 74077
QY 302 rGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyValGlnAsnProGluAsnAsp 322
Db 74078 AGAAGAGATGCTTAAAGATATCTTTCTCTTACATTGGGAAGAAATCCTCGAAACCATTA 74137
QY 322 nPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProAr 342

```

Db	74138	CTTCACTTTACGGAATCCTTTTCTCCACCAATCTCTAATCTCGCTTGGCTCAATCCGG	74197
Qy	342	gAspilleProLeuLeuTyrLeuGlnArgLysille-----	353
Db	74198	CGATATCTCTGCTATACCTCCAGAGAAAGGT-GAGCTTTTTTTCGGGTTTTTGTGATCAT	74256
Qy	354	-----GlnLysAl 356	
Db	74257	TTTAAACGAAAGAGTTTTTCAGCATGTTTTTAATGTTTATTATCATCTCTTAGATTCAAAAGC	74316
Qy	356	aprometGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHi	376
Db	74317	TCAATGGGATCCTTGAAGCAAGCAAGCTCAGAAGAAATGCTTGCAGAAAGAAATCA	74376
Qy	376	sArqLysGlnIleAspGlnSerIleThrAspilleLeuArgLeuSerValLysGlnThrAs	396
Db	74377	TAGGAACAAATCGATCAGACGATTACACACATCTCGCGCTTTCAGTTAAACAAACCAA	74436
Qy	396	nValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspTrpAs	416
Db	74437	TGTCTTAAATCTCTTAACTTCCACAAGAACACAGGACAGCCTCTTGTAGACGATTGGGA	74496
Qy	416	pCysPheLysThrLeu-----	421
Db	74497	TTGCTTCAAGACTCTAGTAACAAACACATCTCAAAACCTTGTTACTGTGTTCTACGCAA	74556
Qy	422	-----ValAsnSerPh 425	
Db	74557	CAACCATTCGATTATTACTAAACACGTGTATATCGAATCGAAATCCAGGTTAATAGCTT	74616
Qy	425	elysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAs	445
Db	74617	CAAGAACTACTCGGGTGCACCGTGCATTCACGATTGAAGTATACAGAGCGCTTGCCAA	74676
Qy	445	nileCysAsnMetGlyValaspValLysGlnThrValSerAlaIleGluAlaCysSe	465
Db	74677	TATCTCAATATGGGAGTGATGTGAAGCAAACTGTTTCAGCCATTGAACAGCTTGTTC	74736
Qy	465	rMet 466	
Db	74737	GATG 74740	

RESULT 8

US-09-803-736-443
 ; Sequence 443, Application US/09803736
 ; GENERAL INFORMATION:
 ; APPLICANT: Bush, David F.
 ; APPLICANT: Levin, Irena M.
 ; APPLICANT: Norris, Susan R.
 ; APPLICANT: Rounsley, Steven D.
 ; APPLICANT: Wiegand, Roger C.
 ; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
 ; FILE REFERENCE: 38-10(15493)D
 ; CURRENT APPLICATION NUMBER: US/09/803,736
 ; CURRENT FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 09/534,859
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 04983.0206CPUS01 38-10
 ; PRIOR FILING DATE: 2000-10-20
 ; NUMBER OF SEQ ID NOS: 1582
 ; SEQ ID NO 443
 ; LENGTH: 83253
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-09-803-736-443

Alignment Scores:
 Pred. No.: 3,26e-205 Length: 83253
 Score: 2188.00 Matches: 463
 Percent Similarity: 67.89% Conservative: 0
 Best Local Similarity: 67.89% Mismatches: 3
 Query Match: 89.16% Indels: 219

DB:	31	Gaps:	7
US-09-934-066-2 (1-466) x US-09-803-736-443 (1-83253)			
Qy	1	MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuLysPhe 20	
Db	72700	ATGCTAGTCCCTCTTGGTCACCTTCAGATCTCTGTTTCTTCATGCTTGTCTTATCTTC	72759
Qy	21	SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnAspValIgluSerSerAspLys 40	
Db	72760	TCAGCTGAGTCGCGCAAAACCCAAATGCTGAAGATAATGATGTTCAATCTAGCGACAAG	72819
Qy	41	SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60	
Db	72820	AGTGCAAAAGGCACACGATGGCTGTTTGTAGTTGCTGGATCAAAATGAATATATAACTAC	72879
Qy	61	ArgHis-----	62
Db	72880	AGCATCAGCTGTTTAAATATGTTTGAACGTTTAAACATAACAAAAAAGAGTCCA	72939
Qy	63	-----GlnAlaAspilleCysHis 68	
Db	72940	ACGAGATTGTTATGAACATAAATCGACCGACGCTTTTATTTCACAGCTGACATATCCAC	72999
Qy	69	AlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMetTyr 88	
Db	73000	GGGTATCAGATCTCCGAAAAGCGGTTTAAAGATGAAACATCATCTGTTTATGTAT	73059
Qy	89	AspAspIleAlaPheSerSerGluAsnProAcqProGlyValIleIleAsnLysProAsp 108	
Db	73060	GATGATATCGGTTTTCTCGGAGAATCCTAGCCCTGGAGTTATCATTAATAAACCCAGAT	73119
Qy	109	GlyGluAspValTyrLysGlyValProLys-----	118
Db	73120	GGAGAAGATGTTTATAAAGAGGTTCCCTAA-GGTTCTTATTCTTCTACTCTCTTCTGCGTTA	73178
Qy	118	-----	118
Db	73179	TTTCTACGTTGAATTCATATATATATATATATATCAAGTTTGTGTTTGTGGGTAG	73238
Qy	119	AspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGlu 138	
Db	73239	GACTACACATAAGAGCTGTTAATGTTCAAACTCTCAATAGTGTACTTGGAAATGAA	73298
Qy	139	SerGlyValThrGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePhe 158	
Db	73299	AGTGGCGTCACAGGAGGAAATGCAAAAGTTGTGAAAGTGGTCTTAATGATAATATCTTC	73358
Qy	159	IleTyrTyrAlaAspHisGlyAlaProGlyLeuIle-----	170
Db	73359	ATCTATTATGCTGACCATGGAGCTCCTGCTTAAATAGGTTTTTCTTAATTTTATGAAATTA	73418
Qy	170	-----	170
Db	73419	TTACGTACCATCAATCCATATCTATAATAAAGATTTTCTCTTGACTACGAAACCGCG	73478
Qy	171	-----AlaMetProThrGlyAspGluValMetAlaLysaspPheAsnGluValLeu 187	
Db	73479	ATTTTCTCAGCATGCCACTGGTGATGAAGTTATGCAAAAGATTTCAATGAAGCTTTC	73538
Qy	188	GluLysMetHisLysArgLysLysTyrAsnLysMetVal-----	200
Db	73539	GAGAAGATGCATAAGAAAAAATAACAAAGATGTT-ATATACTCAACCATCTGTTA	73597
Qy	200	-----	200
Db	73598	CCTAGCTTTATACATATGTTCTGTTTGTGAATCTCTATGGTGTGTTTTTTTGGATGTT	73657
Qy	201	-----IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLys 218	
Db	73658	TAGGTGATCTATGTTGAACCATGTGAATCAGGAAGTATCTTCAAGGGATTTAAAGAAA	73717
Qy	219	AsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyr 238	

QY 182 AspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValle 201
 DB 547 GATCTGATTCAGTCTTGAAGAAAGACGCTCTTCTGGAAGTTATAAAAGCCTAGTATTT 606
 QY 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyLeuLeuLysLysAsnLeuAsn 221
 DB 607 TATCTAGAGCATGTGAATCTGGAGTATCTTTGAAGGCTCTTCTCTGAAGGCTCTGAAT 666
 QY 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241
 DB 667 ATCTATGCAACACAGCTTCAATGAGAAAGCAGTTGGGGAACATATTTCTCTGGG 726
 QY 242 SerTyrProProProSerGluLeuGlyThrCysLeuGlyAspThrPheSerIleSer 261
 DB 727 GAGTATCTAGTCTCTCCCTGTAATGAAACATGCTGGTGACCTGACAGTGTGCT 786
 QY 262 TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHis 281
 DB 787 TGGATGGAGACAGTGCATACATACAAATTTGCAACAGAACTTTACATCAACAAATACGAA 846
 QY 282 ValValLysArgArgVal-----GlySerAspValProGluThrSerHisValCysArg 299
 DB 847 TTGCTCAACAAAGGACTATGAAATGCAATTTCAATTTATGCTTCCACAGTGTGCTGAG 903
 QY 300 PheGlyThrGluLysMetLysAspTyrLeuSerSerTyrIleGlyArgAsnProGlu 319
 DB 904 TATGGTGACATAGGGCTTACCGGAGAACTCTCTCTTATATTTGGGTACAAATCTCTGCT 963
 QY 320 AsnAspAspPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuVal 339
 DB 964 AATGATANTTTACTTTTGTGCTTAA--AACTCATTTGGTGCCACTTCAAAAGCAGTC 1020
 QY 340 AsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGly 359
 DB 1021 AACCAACGTGATCGATCTCATCTTATTTTGGGATAAGTTCCGCAAGCTCTCTGGGT 1080
 QY 360 SerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGln 379
 DB 1081 TCTTCTAGGAAAGCTCGAGCTGAGAACAAATTTTGAAGCAATGCTCACAGAAATGCAT 1140
 QY 380 IleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsn 399
 DB 1141 ATAGATGACAGCATGAACGTATTTGAAAGCTCTTCTTTGGCATTGAAAGGCTCCAGAA 1200
 QY 400 LeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLys 419
 DB 1201 CTGCTTAGCAGCTTAGACCTGCTGGCAACCACTTGTTCATGACTGGGACTGCCCTTAA 1260
 QY 420 ThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyr 439
 DB 1261 ACATTGGTTAGGACTTTTGAGACACATTTGGATCCCTGCTCAGTATGGGATGAAACAT 1320
 QY 440 ThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAla 459
 DB 1321 ATGAGGCTCTTCCAAACITCTCCACGCTGGATACGAAAGCAATGGCTGAGGCC 1380
 QY 460 IleGluGlnAlaCys 464
 DB 1381 TCAGCACAGCATGT 1395

RESULT 10

US-60-312-544-89
 : Sequence 89, Application US/60312544
 : GENERAL INFORMATION:
 : APPLICANT: Cao, Yongwei
 : APPLICANT: Edgerton, Michael D
 : APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Kovalic, David K.
 : APPLICANT: Liu, Jingdong
 : APPLICANT: Stein, Joshua
 : TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
 : FILE REFERENCE: 38-10(52726)A
 : CURRENT APPLICATION NUMBER: US/60/312,544

: CURRENT FILING DATE: 2001-08-15
 : NUMBER OF SEQ ID NOS: 10730
 : SEQ ID NO 89
 : LENGTH: 1758
 : TYPE: DNA
 : ORGANISM: Zea mays
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: (1)..(1449)
 : OTHER INFORMATION: Clone ID: 7000756373_FLI
 : US-60-312-544-89

Alignment Scores:
 Pred. No.: 1,66e-117 Length: 1758
 Score: 1287.00 Matches: 244
 Percent Similarity: 68.82% Conservatives: 76
 Best Local Similarity: 52.47% Mismatches: 135
 Query Match: 52.44% Indels: 10
 DB: 75 Gaps: 4

US-09-934-066-2 (1-466) x US-60-312-544-89 (1-1758)

QY 8 PheGlnIleLeuValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr 27
 DB 7 TTCCGATCTCTTCTTCCTCCGCCACCTCATCCCTCGCTCCGTCGCCCGCCACGAT 66
 QY 28 GlnLeuLeuAsnAspAsnValGluSer-----SerAspLysSer 41
 DB 67 ATTCTCCGTTTACCTCCGAAAGCATCCACTTTTTCAAAGCACCCGTCGCGATCAAAAC 126
 QY 42 AlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrAsnTyrArg 61
 DB 127 GATGAGGCGACGAGTGGCGCTTTTAATTCGCGTTTCCAATGGCTACTGGAATTACAG 186
 QY 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGlu 81
 DB 187 CACCATCTGATGTTTCCCATGCTATCACTACTGAGAAAGTGGTCTCAAGAAGAA 246
 QY 82 AsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGly 101
 DB 247 AATATTGTTGTTATTTATGATGATGATGATGCTTTCACCAAGACAGAACCCGACCTGGA 306
 QY 102 ValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThr 121
 DB 307 GTCAATTATTAAACAGTCCACATGGAAATGATGTTTACAAGGAGTCCCTTAAGGATTACAT 366
 QY 122 LysGluAlaValAsnValGlnAsnPheTyrAsnValLeuGlyAsnGluSerGlyVal 141
 DB 367 GGTGAAGATGTAAGTGTGGCAACTTTTTCGCTATCTTGGAAATAAGTCAGCTCTT 426
 QY 142 ThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyr 161
 DB 427 ACTGGTGGCAGTGGGAGGTTGGATAGTGTGCCAATGATCATATATATATATATAC 486
 QY 162 AlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLys 181
 DB 487 TCTGATCATGGCGCTCTGGAGTGTAGGATCCCTACTAATCCATACATGATGTCATCT 546
 QY 182 AspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValle 201
 DB 547 GATCTGATGCAAGTCTTGAAGAAGACGATGCTTCTGGAAGTTATAAAAGCCTAGTATTT 606
 QY 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
 DB 607 TATCTAGAGCATGTGAATCTGGGAGTATCTTTGAAGGCTCTTCTCTGGAAGGCTCTCAAT 666
 QY 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241
 DB 667 ATCTATGCAACACAGCTTCAATGAGAAAGCAGTTGGGGAACATATTTCTCTGGG 726
 QY 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261
 DB 727 GAGTATCTAGTCTCTCTCTGATGATGAAACCTGCTGCTGCTGACCTGACAGTGTGCT 786

Db	111	TC	TCTCTCCAATCTCACACCCATCATGACCGGTTTCCGGATCCTCTTTTCTCGTCGGCACCC	170
Qy	13	Phe	Leu-----HISAlaLeuLeuIlePheSerAlaIleuSerArg	25
Db	171	TT	CATACCCCTCGCCTCGGTGCGCGCCAGCATATTCTCGGTACCTCCGAAGCTTCC	230
Qy	26	Lys	ThrGlnLeuLeuAsnAspAsnAspValGluSerSerAspLysSerAlaLysGlyThr	45
Db	231	AG	GTCTCTCAAGCACCTGCTAAT-----GCCGATCAAAACGATGAGGCCAC	278
Qy	46	Arg	TrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTrpArgHis-GlnAlaAs	65
Db	279	AG	TGGGCCGCTTTAGTTGCGGTCCCATGGCTACTGGAAATTCAGAGCACCCAGTCTGA	338
Qy	65	P	IleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleLeuVa	85
Db	339	TG	TTTGCCATGCGGTATCACTACTGAGAAAGGTGGTCTCAAGAAGAAATAATTCTGT	398
Qy	85	I	peMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAs	105
Db	399	AT	TATGTATGATGATGACATTCCTTTCAACGAAGAGAACCCCGACCTGGAGTCAATTATTA	458
Qy	105	N	ysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaVa	125
Db	459	CAG	TCCACATCGAAATCATCTTTACAAGGAGTCCCTAAGGATTACATTGGTCAAGATGT	518
Qy	125	Las	nValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGlyGlyAs	145
Db	519	AA	CTGTGGCACTTTTCTGCTATATCTTGGAAATAAGTCAGCTCTTACTGGTGGCAG	578
Qy	145	n	GlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisG	165
Db	579	TG	GAAAGTTGTGGATAGTGGCCCAATGATCATATATTTATATCTACTCTGATCATGG	638
Qy	165	Y	AlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnG	185
Db	639	CG	TGGCGGAGTCTAGGGATGGCTAATCATCATCATATGATGATGCCCATCTGATGA	698
Qy	185	u	ValLeuGluLysMethisLysArgLysLysTyrAsnLysMetValIleTyrValGluAl	205
Db	699	AG	CTTTGAAGAAAGAACCATCTCTCGAACCTTATAAAGCCTTAGTATTTTATCTAGAGC	758
Qy	205	a	CysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaVa	225
Db	759	AT	GTAATCTCGGAGTATCTTTCAAGGTCTTCTCCAGAAAGGTCTGAATATCTATGCAAC	818
Qy	225	I	ThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProPr	245
Db	819	AA	CAGCTTCAAAATGCTCAAGAAAGCAGTTTGGCGAACATATTGTCTGGGAGTATCCTAG	878
Qy	245	o	ProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAs	265
Db	879	TC	TCCCTCCCTGTAATATCAAAACCTGCCTGGCTGACCTCTACAGTGTGCTGGATGGAAGA	938
Qy	265	p	SerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHisValValLysAr	285
Db	939	TAG	TGACATACAAATTTGGCAAGAACATTTACATCAACAATACGATTTGGTCCAAAGA	998
Qy	285	g	ArgVal-----GlySerAspValProGluThrSerHisValCysArgPheGlyThrG	303
Db	999	AAG	CATATGAATGGAAATTCATATC---TATGGTCTCCAGCTGATCAGTATGCTGATCAT	1055
Qy	303	u	LysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPh	323
Db	1056	AG	GCGCTTAGCAAGAACAAATCTCTCTATATTTTGGGTACAAATTCCTGCTAATCATATT	1115
Qy	323	e	ThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProArgAs	343
Db	1116	TAC	TTTTTGTCTTAAACACTCATATG---GTGCCACCTTCAAAAGCAGTCAACCAAGTGA	1172
Qy	343	p	IleProLeuTyrLeuGlnArgLysIleGlnLysAlaProMetCysIleSerLeuGluSe	363

•

```

Db      1173  TGCAATCTCATCCATTCTGGGATTAAGTTCCCGAAAGCTCCCTGGTGGTCTCTTAGGAA 1232
Qy      363  rLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSe 383
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1233  AGCTGCAGCTGAGAAAGAAATCTGGAAGCAATGCTCTCAGAGATGATATAGATGACAA 1292
Qy      383  rIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSe 403
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1293  CATGAAACTTATTGGAAGCTCTTATTGGCATTTGAAAGGGTCCAGAACTGCTTAGCAG 1352
Qy      403  rThrArgThrThrGlnProLeuValAspTTPAspCysPheLysThrLysThrValAs 423
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1353  TGTAGACCTGCTGGGCAACCACTGTTGATGACTGGGACTGCTTAAACATTTGGTTAG 1412
Qy      423  nSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLe 443
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1413  GACTTTTGAGACACATTTGTCATGCTCTGATGATGGAAGGATGAAACATATGAGGTCCT 1472
Qy      443  uAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluAla 463
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1473  TGCAAACTTCTGCAACGCTGGAATACGAAAGCAAAATGGCTGAGGCTTCAGCACAA 1532
Qy      463  aCys 464
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1533  ATGT 1536

RESULT 12
US-09-684-016-3227
; Sequence 3227, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684,016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 3227
; LENGTH: 1914
; TYPE: DNA
; ORGANISM: Glycine max
US-09-684-016-3227

Alignment Scores:
Pred. No.: 9.21e-115 Length: 1914
Score: 1260.00 Matches: 247
Percent Similarity: 66.39% Conservative: 73
Best Local Similarity: 51.24% Mismatches: 137
Query Match: 51.34% Indels: 25
DB: 27 Gaps: 6

US-09-934-066-2 (1-466) x US-09-684-016-3227 (1-1914)
Qy      2  SerSerProLeuGlyHis-----PheGlnIleLeuVal 12
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      111  TCTTCCAAATCTCACACCCATCATGACCGGTTCCGATCCTCTTCGTCGGCACCC 170
Qy      13  PheLeu-----HisAlaLeuIlePheSerAlaGluSerArg 25
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      171  TTATCAACCTCGCTCCGTCGGTCCCGCCAGCATATCTCCGGTATACCTCCGAAGCTCC 230
Qy      26  LysThrGlnLeuLeuAsnAspValIleuSerSerAspLysSerAlaLysGlyThr 45
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      231  AGGTTCTTCAACACCTGCTAAT-----GCCATCAACAGATGAGGCACC 278
Qy      46  ArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHis-GlnAlaAs 65
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      279  AGGTGGGCGGCTTTAGTTGCGGTTCCCAATGGCTACTGGAATACAGGCACCCAGTCTGA 338
Qy      65  pIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleLeva 85
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

Db      339  TGTTTGCCATCGCATCAACTACTAGGAAAGTGGTCTCAAAAGAAATAATATTGTTGT 398
Qy      85  lPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAs 105
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      399  ATTATGTATGATGACATTCCTTTCAACGAAGAGAACCCCGACCTGGAGTCATTATTA 458
Qy      105  nLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaVa 125
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      459  CAGTCCACATGGAAGATGATGTTTACAAGGGAGTCCCTAAGGATTACATTGCTGAAG 518
Qy      125  lAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerClyValThrGlyGlyAs 145
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      519  AACTGTTGGCACTTTTTCCTCTATCTATGGAATAAGTCAGCTCTTACTGGTGCGAC 578
Qy      145  nGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisG 165
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      579  TGGGAAGTTGTGATAGTGGCCCCCAATGATCATATATTTATATACTACTCTCATCG 638
Qy      165  yAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnG 185
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      639  CGTCCGGAGTGTAGGATGCTACTAATCCATACATGATGATGATGATGATGATGATGA 698
Qy      185  uValLeuGluLysMetHisLysLysTyrAsnLysMetValIleTyrValGluAl 205
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      699  AGTCTTGAAGAAGACATGCTTCTGGAACCTTATAAAGCCTAGTATTTCTAGAGGC 758
Qy      205  aCysGluSerGlySerMetPheGlyIleLeuLysLysAsnLeuAsnIleTyrAlaVa 225
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      759  ATCTGAATCTGGCAGTATCTTTCAAGGCTCTCTCCAGAAGGCTGTAATATCTATGCA 818
Qy      225  lThrAlaAlaAsnSerLysGluSerSerTyrGlyValTyrCysProGluSerTyrProPr 245
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      819  AACAGCTTCAATTCCTGAAGAAAGACAGTTGGGAACATATTTGCTCGGGAGTATCCTAG 878
Qy      245  oProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerThrLeuGluAs 265
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      879  TCCTCCCCCTGAATATGAAACCTGCGCTGGGTGACCTGACAGTGTGCTGGATGGAAGA 938
Qy      265  pSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysAr 285
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      939  TAGTGACATACACAATTTGGGACAGAAACATTTTACATCAACAATACGACTTGGTCAA 998
Qy      285  gArgVal-----GlySerAspValProGluThrSerHisValCysArgPheGlyThrG 303
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      999  AAGGACTATGAATGGAAATTCATC---TATGGTTCCACGCTGATGACGATGCTGCAT 1055
Qy      303  uLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPh 323
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1056  AGGCTTAGCAAGAACAACTCTCTCTTATATTTGGGTACAAATCTCTGCTAATGATAAT 1115
Qy      323  eThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProArgAs 343
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1116  TACTTTTGTGCTTAAACACTCATTTG---GTGCCACCTTCAAAAGCAGCTCAACCACTGA 1172
Qy      343  pIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetClySerLeuGluSe 363
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1173  TGCAGATCTCATCCATTTCTGGGTAAGTTCCGCAAGCTCCTGCTGGGTCTTCTTAGGA 1232
Qy      363  rLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSe 383
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1233  AGCTGAGCTGGAAGAAATCTTGGAAAGCAATGCTCTCACAGAAATGATATAGATGACAA 1292
Qy      383  rIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSe 403
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1293  CATGAAACTTATTGGAAGCTCTTATTGGCATTTGAAAGGGTCCAGAACTGCTTAGCAG 1352
Qy      403  rThrArgThrThrGlyGlnProLeuValAspTTPAspCysPheLysThrLysThrLeuVal 423
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1353  TGTAGACCTGCTGGGCAACCACTGTTGATGACTGGGACTGCTTAAACATTTGGTTAG 1412
Qy      423  nSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLe 443
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      1413  GACTTTTGAGACACATTTGTTGGATCCCTCTGCTCAGATGGAATGGAACATATAGAGGTCCTTT 1472

```


QY 443 uAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGlnAl 453
 D6 1473 TGCAAACTTCTGCAACGCTGGAAATACGAAAAGAGCAATGGCTGAGGCCTACGACACAGC 1532

Qy 463 aCys 464
|||
Db 1533 ATGT 1536

```

RESULT 13
US-09-513-996A-37409
; Sequence 37409, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

```

```

; TITLE OF INVENTION: ENCODED THERESY
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513.996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 37409
; LENGTH: 1485
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..1485
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..1485 / Ceres Seq. ID 1821917
; US-09-513-996A-37409

Alignment Scores:
Pred. No.:
Score: 1.64e-113 Length: 1485
Matches: 233
Conservative: 71
Mismatch: 121
Indels: 10
Gaps: 4
DB:

US-09-934-066-2 (1-466) x US-09-513-996A-37409 (1-1485)

```

Qy	196	TyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIle	215
Db	625	TATAAAAGCCTGGGTGTTTTTTCGGAAGCTTCGGAATCTGGAAGTATCTCTTTTGAAGGCGCTT	684
Qy	216	LeuLysLysAsnLeuAsnIleTyrAlaValAlaAlaAsnSerLysGluSerSerTrp	235
Db	685	CTTCTCTGAGGGTTTTGAACATCTATGCCAACACTGCATCAACCGCGAAGAACAGTGG	744
Qy	236	GlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGly	255
Db	745	GCTACCTATTGCCCTCGAGAGGAACACGCTCTCCACCGGAGCTATCAAACTTGTGTAGGT	804
Qy	256	AspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThr	275
Db	805	GACTTGTACAGTGTTCCTGGATGGAGATAGTGGTATGCCACAATTTACAGACTGAGACT	864
Qy	276	LeuGluGlnGlnTyrHisValValLysArgArg-----ValGlySerAspValpro	292
Db	865	CTGCACGACCAATATGAACCTTGGAAGAAGGAGCTGCACCTGTTGGGTACTCTTAT---	921
Qy	293	GluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLysSerSer	312
Db	922	---GGTCTCTCATGTCACCAATATGGCGATGTAGGAATTTAGCAAGCATTAATCTCGATCTT	978
Qy	313	TyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGlu-----SerPheSer	330
Db	979	TATATGGGAACAAACCCCTGCCAATGACAAATTTTACCTTTTCGGGATTCGCAATTCACTAAG	1038
Qy	331	SerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLysGln	350
Db	1039	CCACCT-----TCAAGAGTTTACAACACCGCTGATGCAGATCTTGTTCATTTTGG	1089
Qy	351	ArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu	370
Db	1090	GAAGAAGTACCAGAAAGCACCAGAAGCTTCAGCAAGAAAAACAAGAGCTCACAACCAAGTA	1149
Qy	371	LeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeu	390
Db	1150	CTTGAAGCCATGCTCTCACAGACTTCATATTGACATAGCGTGCTACTCGTCGGAANAATC	1209
Qy	391	SerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnPro	410
Db	1210	TTGTTTGGCATTTTCGACAGGTCCTGAAGTCTTAACAAGATPACGGTCTGCTGGCAACCT	1269
Qy	411	LeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGly	430
Db	1270	CTPATGCGATGACTCGAAGTCTGCTTAAAAATACAGGTGAGCTTTTCGAGAGGCGACTGTGGA	1329
Qy	431	AlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGly	450
Db	1330	TCGCTGTCTCAGTACGGTATCAAGACATGAGGTCTTTTGCAAACATCTGCAATGCAGGG	1389
Qy	451	ValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSer	465
Db	1390	ATTCAAAATGGAGCAAAATGGAGGAGGAGGCTTTCACAGGCTTGTACC	1434

```

RESULT 14
US-09-708-427-10759
: Sequence 10759, Application US/09708427
: GENERAL INFORMATION:
: APPLICANT: N. ALEXANDROV et al.
: TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
: TITLE OF INVENTION: THEREBY
: FILE REFERENCE: 2750-1243P
: CURRENT APPLICATION NUMBER: US/09/708,427
: CURRENT FILING DATE: 2000-11-09
: NUMBER OF SEQ ID NOS: 85364
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 10759
: LENGTH: 1485
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana

```


FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: 1..1485
 : OTHER INFORMATION: any n = a, g, c, t, unknown, or other
 : NAME/KEY: misc_feature
 : LOCATION: 1..1485
 : OTHER INFORMATION: Ceres Seq. ID 1821917
 : US-09-708-427-10759

Alignment Scores:

Pred. No.: 1.64e-113 Length: 1485
 Score: 1246.00 Matches: 233
 Percent Similarity: 69.89% Conservative: 71
 Best Local Similarity: 53.56% Mismatches: 121
 Query Match: 50.77% Indels: 10
 DB: 28 Gaps: 4

US-09-934-066-2 (1-466) x US-09-708-427-10759 (1-1485)

Qy 36 GluSerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsn 55
 Db 145 GAAACGACGACGATTAACCTCCGCTACTAGTGGCTCTTCTAGTCGCGGATCAGC 204
 Qy 56 GluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLys 75
 Db 205 GGATATTGGAATTACAGGCATCAGGCTGATATATGCCATGCCCTATCAACTTCTGAGGAA 264
 Qy 76 GlyGlyLeuLysAspGluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSer 95
 Db 265 GTTGAGATTGAAAGAGAGAAATATTGGTATTCTATGATGATGATATGCTACAAATTAC 324
 Qy 96 GluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGly 115
 Db 325 GAGAACTCAAGGCCTGGAACCACTTATCAACAGCCCTCATGGAAGAGATGCTTATCAAGGA 384
 Qy 116 ValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLys 135
 Db 385 GTTCCCAAGGATTATCTAGGATGATGTCATGTAATCTATTGCTGTGATGCTCTT 444
 Qy 136 GlyAsnGluSerGlyValThrGlyValAsnGlyLysValValLysSerGlyProAsnAsp 155
 Db 445 GGAGACAAACTCCTGTAAGGGGGAAGTGGAGGTTCTGGTAGTGGTCTATATGAT 504
 Qy 156 AsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
 Db 505 CATATCTTCATATCTACAGTACCACCTGGTGGTCTGGAGTTCTTGGGATGCCAACTTCT 564
 Qy 176 AspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLys 195
 Db 565 CTTTACCTATPATGCAATGATCAATGATGCTTGAAGAAGAAACATGCTTTAGGAACA 624
 Qy 196 TyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIle 215
 Db 625 TATAAAGCTTGTTGTTTATCTCGAAGCTTGGCAATCTGGAAGTATCTTTGAAGGCTT 684
 Qy 216 LeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrp 235
 Db 685 CTTCTCGAGGTTTGAACATCTATGCCAACATGATCAACCCGCGGAAAGACAGTTGG 744
 Qy 236 GlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGly 255
 Db 745 GGTAACCTATTGCCCTGGAGGAGAACCCAGTCCCTCCACCGGAGTATGAAACTTCTTAGGT 804
 Qy 256 AspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThr 275
 Db 805 GACTTGTACAGTGTCTGCTGGATGGAGATAGTGGTATGACCAATTTACAGACTGAGACT 864
 Qy 276 LeuGluGlnGlnTyrHisValValLysArgArg-----ValGlySerAspValPro 292
 Db 865 CTCACACGCAATATGAACCTTGTGAAAGAGGAGGACTGCACCTCTGGTACTCTTAT--- 921
 Qy 293 GluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSer 312
 Db 921 GATATTTGGAATTACAGGCATAGGCTGATATATGCAATGCCATCACTATCACTTCTGAGGAA 264

Db 922 ---GGTTCTCATGTCATGCAATATGCGGATGTAGGAATTAGCAAGGATAATCTCGATCTT 978
 Qy 313 TyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGlu-----SerPheSer 330
 Db 979 TATATGGGAACAACCCCTGCCAATGACAATTTACCTTTGCGGATCGGAATTCACATAAG 1038
 Qy 331 SerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGln 350
 Db 1039 CCACCT-----TCAAGAGTTACAACACCGCTGATGCAGATCTTGTTCATTTTGG 1089
 Qy 351 ArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu 370
 Db 1090 GAAAGTACCGAAAGCACCAGAGGTTCCAGCAAGAAACACAGAGCTCAGAAGCAAGTA 1149
 Qy 371 LeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeu 390
 Db 1150 CTTGAAGCCATGCTCACAGACTTCATATTGACAATAGCGTGATCTCTCGCAAAATC 1209
 Qy 391 SerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnPro 410
 Db 1210 TTGTTTGGCATTTTCGAGAGGTCCTGAAGTCTTAACAAAGTAGCGTCTGCTGGCAACCT 1269
 Qy 411 LeuValAspAspThrAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGly 430
 Db 1270 CTATGTCGATGACTGGAACTGCCCTTAAATAATCAGGTGAGAGCTTTCGAGAGGCAC 1329
 Qy 431 AlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGly 450
 Db 1330 TCCTGCTCTCAGTACGGTATCAAGCACATCAGGTCTTTTGCACAACTCTCCAATGCAGG 1389
 Qy 451 ValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSer 465
 Db 1390 ATTCAATGAGCAAAATGGAGGAGGAGGCTTTCAGAGCTTGTACC 1434

RESULT 15

US-09-935-625-11673
 : Sequence 11673, Application US/09935625
 : GENERAL INFORMATION:

APPLICANT: N. ALEXANDROV et al.
 : TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPA
 : TITLE OF INVENTION: MODULATING VARIOUS RESPONSES
 : FILE REFERENCE: 2750-1481P
 : CURRENT APPLICATION NUMBER: US/09/935,625
 : CURRENT FILING DATE: 2001-08-24
 : NUMBER OF SEQ ID NOS: 33136
 : SEQ ID NO 11673
 : LENGTH: 1485
 : TYPE: DNA
 : ORGANISM: Arabidopsis thaliana
 : FEATURE:
 : NAME/KEY: Misc_feature
 : LOCATION: 1..1485
 : OTHER INFORMATION: Ceres Seq. ID no. 1821917
 : US-09-935-625-11673

Alignment Scores:

Pred. No.: 1.64e-113 Length: 1485
 Score: 1246.00 Matches: 233
 Percent Similarity: 69.89% Conservative: 71
 Best Local Similarity: 53.56% Mismatches: 121
 Query Match: 50.77% Indels: 10
 DB: 35 Gaps: 4

US-09-934-066-2 (1-466) x US-09-935-625-11673 (1-1485)

Qy 36 GluSerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsn 55
 Db 145 GAAACGACGACGATTAACCTCCGCTACTAGTGGCTCTTCTAGTCGCGGATCAGC 204
 Qy 56 GluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLys 75
 Db 205 GGATATTGGAATTACAGGCATAGGCTGATATATGCAATGCCATCACTATCACTTCTGAGGAA 264

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 18:27:24 ; Search time 691 Seconds
(without alignments)
3732.872 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHQILVFLHALLIF.....CNGVDVKQTVSAIEQACSM 466

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 6653949 seqs, 2767611981 residues

Total number of hits satisfying chosen parameters: 13307898

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
-O/cgn2_1/USPTO.spool/US0934066/runat_19052003_163701_24433/app_query.fasta_1.647
-DB=Pending Patents.NA.New -OFMT=fastap -SUFFIX=p2n.rnnp -MINMATCH=0.1
-LOOPCI=0 -LOOPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US0934066.rcgn_1_1_327@runat_19052003_163701_24433
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT
-DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending_Patents_NA_New.*

1:	/cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
2:	/cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3:	/cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4:	/cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5:	/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6:	/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq2.*
7:	/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq3.*
8:	/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
9:	/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq2.*
10:	/cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*
11:	/cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1287.5	52.4	1758	US-10-425-114-8500	Sequence 8500, Ap
2	1284.5	52.3	1937	US-10-424-599-76825	Sequence 76825, A
3	1279.5	52.1	2133	US-10-424-599-59168	Sequence 59168, A
4	1273	51.9	2006	US-10-424-599-59165	Sequence 59165, A
5	1205.5	49.1	1881	US-10-425-114-28946	Sequence 28946, A
6	1204.5	49.1	1786	US-10-425-114-35367	Sequence 35367, A
7	1204.5	49.1	1794	US-10-425-114-30424	Sequence 30424, A

ALIGNMENTS

RESULT 1

US-10-425-114-8500

; Sequence 8500, Application US/10425114

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; SEQ ID NO 8500

; LENGTH: 1758

; TYPE: DNA

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: 700756373_FLI

; US-10-425-114-8500

Alignment Scores:

Pred. No.: 3.27e-121

Score: 1287.00

Percent Similarity: 68.82%

Best Local Similarity: 52.47%

Query Match: 52.44%

DB: 8

Length: 1758

Matches: 244

Conservative: 76

Mismatches: 135

Indels: 10

Gaps: 4

Sequence 28422, A
Sequence 15845, A
Sequence 15458, A
Sequence 34046, A
Sequence 27715, A
Sequence 35501, A
Sequence 49438, A
Sequence 10538, A
Sequence 19102, A
Sequence 6839, Ap
Sequence 15848, A
Sequence 30, Appl
Sequence 21292, A
Sequence 452, App
Sequence 3, Appl1
Sequence 6167, Ap
Sequence 33, Appl
Sequence 1847, Ap
Sequence 4031, Ap
Sequence 4031, Ap
Sequence 114, App
Sequence 32083, A
Sequence 32083, A
Sequence 32078, A
Sequence 32084, A
Sequence 32084, A
Sequence 32084, A
Sequence 32079, A
Sequence 32079, A
Sequence 71, Appl
Sequence 446, App
Sequence 9493, Ap
Sequence 30081, A
Sequence 38671, A

Qy	2	SerSerPro---LeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe	20
Db	148	TCACACCATCATCATGACCGTTTTCGAGTCTCTTCTTCCTCGCACCATCATCACCTC	207
Qy	21	SerAlaGluSerArgIlyThrGlnLeuLeuAsnAspAsnAspValGlu---	36
Db	208	GGCTCCGGTCCCGCCACCATATTCTCCGGTTACCTCCGAAGCATCCACATTTTTCAAA	267
Qy	37	-----SerSerAspLysSerAlaLysGlyThrArgTIpAlaValLeuValaGly	53
Db	268	GCACCTGCCTAATCCGCATCAAAACCATGAGGCACAGTGGCGGTTTACTTCCGGT	327
Qy	54	SerAsnGluTyTyTyrAsnTyArgHisGlnAlaAspIleCysHisAlaTyArgIleLeu	73
Db	328	TCCAATGGCTACTGGAATTACAGGCACAGTCTGATGTTGCCATGCATATCAACTACTG	387

Db	508	: :	CATGAGGCTCTGGGTCCTGGGATGCTGCTGCTTACTTATACGGGATGATCTG	567
Qy	184	: :	AsnGluValLeuGluLysMetHisLysArgLysLysTyAsnLysMetValIleTyrVal	203
Db	568	ATTGAAGTCTTTGAAGAAAACGATGCTCTGGAACATATAAAACCTAGTATTTTATCTG	627	
Qy	204	GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr	223	
Db	628	GAGGCATGTGAATCTGGGAGTATCTTTGAAGGTCCTTCTCCTGAAGATATCAAAATTTAT	687	
Qy	224	AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr	243	
Db	688	GCACCACTGCTTCCAATGCAGAGAAGTAGTTGGGGAACATATATGCCCCGGGAGCTAT	747	
Qy	244	ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu	263	
Db	748	CCTAGCTCTCCCCAGAAATATACAACCTGTCTTGGTGACTTGTACAGTGTGTCTTGGATG	807	
Qy	264	GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal	283	
Db	808	GAAGACAGTGACACACAAATTTGCCAGACAAACTCTGCACCAACAATATAAAATGGTT	867	
Qy	284	LysAlaArg---ValGlySerAspValProGluThrSerHisValCysArgPheGlyThr	302	
Db	868	AAAGAGAGGACTATATCTGGAGATTTCATCTATGCTGCTCACGTGATCGAGTGGTGCAT	927	
Qy	303	GluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsn	322	
Db	928	GTAGGGCTTAGCAGAGATGTTCTCTCCCATTTATTTGGGTACAGATCCTGCTAATGATAAT	987	
Qy	323	PheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProArg	342	
Db	988	TTCACTTTTGGGATGAA--ACTCTTATGGTCACCTTCAAACCAAGTACACACGT	1044	
Qy	343	AspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGlu	362	
Db	1045	GATGCTGATCTATCTCAATTTTGGGATATGTCGCCGAAGCTCTGAGGGTCTCTCAGG	1104	
Qy	363	SerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGln	382	
Db	1105	AAAAATACAGCTCAGAACAAGTTTTTGGAAACAATGCTCACAGAATGCATGTAGACAC	1164	
Qy	383	SerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThr	402	
Db	1165	ACTGTAAACCTGATTTGGGAAGCTTTATTTGGCATTTGAAAGGGTCCAGAAAGTACTCAAC	1224	
Qy	403	SerThrArgThrGlyGlnProLeuValAspAspTrpAspCysPheLysThrLeuVal	422	
Db	1225	GCTGTTAGACCGGTGATCGGCACATGTTGATGACTGCGACTGCTCGAAACCATGGTG	1284	
Qy	423	AsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAla	442	
Db	1285	AGGACTTTTGAGACACATCTGGATCCTTGTCTCAATACGGGATGAACACATGAGGTCC	1344	
Qy	443	LeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGln	462	
Db	1345	TTTGCAAACATCTGCAATGTAGGATAAAGAATGAACAAATGGCTGAGGCTTCAGCACA	1404	
Qy	463	AlaCys 464		
Db	1405	GCTTCT 1410		

```

: RESULT 4
: US 10 424 598-59165
: 59165 Application US/10424599
: GENERAL INFORMATION:
: INVENTOR: LA ROSA THOMAS J
: APPLICANT: KOVALIC DAVID K
: APPLICANT: ZHOU YIHUA
: APPLICANT: CAO YONGWEI
: TITLE OF INVENTION: SOY NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT

```

```

; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 59165
; LENGTH: 2006
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_24438C.1
US-10-424-599-59165

```

Alignment Scores:		
Pred. No.:	1.07e-119	Length:
Score:	1273.00	Matches:
Percent Similarity:	71.03%	Conservative:
Best Local Similarity:	55.14%	Mismatches:
Query Match:	51.87%	Indels:
DB:	8	Gaps:
		2006

US-09-934-066-2 (1-466) x US-10-424-599-59165 (1-2006)

QY	38	SerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyr	57
Db	122	ACAGTACAAACGTCACGGGACCGCTGGCCGCTCTCTCGCCGGTCCCAATGGCGTAC	181
QY	58	TyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGly	77
Db	182	TGGAATTACAGACATCAGCGCTGATGTTTGTTCACGCCATCAAAATPACPAAGGAAAGTGGC	241
QY	78	LeuLysAspGluAsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsn	97
Db	242	CTCAAGAGAAGAAATATTATTCTTTTATGATGATGACATTCGATTCAATGGGGGAAAC	301
QY	98	ProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValPro	117
Db	302	CCAAGCGCTGGAGTCATCATTAACAAACCAAGATGGAGTGATGTATTGAAGGAGTTCCA	361
QY	118	LysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsn	137
Db	362	AAGGATTACACCGCGAAGATGTTACTGTGTGGTAACTTTTTTCGTCTTTACTTTGGAAAC	421
QY	138	GluSerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIle	157
Db	422	AAGTCGGACCTTACTTGTGTGGCAGTGGCAAGCTGTGGACAGCTGGACCTGATGATCATATA	481
QY	158	PheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGlu	177
Db	482	TTTGTATACTATPACTGACCATGGAGGTCCAGGGGTGCTCGGGATGCCCTGCTCTTAC	541
QY	178	ValMetAlaLysAspPheAsnGluValLeuGluLysMethIleLysArgLysLysTyrAsn	197
Db	542	TTATATCGGGATGATCTGATTCAAGTCTTTGAAGAAAAAACAATGCTTCGGAACCTATAA	601
QY	198	LysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLys	217
Db	602	AACCTAGTATTTATCTCGAGCGCATGTCAAATCTCGGAGTATCTTTGAAGCTCTTCTTCCC	661
QY	218	LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyVal	237
Db	662	GAAGATATCAATATTATTGCAACCAACACAGCTTCCAATGCAGAAGAAAGTACTTGGGGAAAC	721
QY	238	TyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThr	257
Db	722	TATTCCTCTGGGGAGTATCTTAGTCTTCCCTCCCCCAAAATATTCAACCTGCTTGGGTGATCTG	781
QY	258	PheSerIleSerTrpLeuGluAspSerAspLeuIleAspMetSerLysGluThrLeuGlu	277
Db	782	TACAGTGTGCTTGGATGGAAGAGTGCACAGACACAAATTTACGAACAGAAACTCTGCAT	841
QY	278	GlnGlnTyrHisValValLysArgArg--ValGlySerAspValProGluThrSerHis	296
Db	842	CAACATATAAATGTTTAAGACAGAGACTATATCTCGAGATTCATACTAGGCTCTCAC	901


```
Db 215 GCGTCGGGACCGGTCGTCGTCATCGCGGCTCCACGGCTACTACAACTACCGC 274
Qy 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGlu 81
Db 275 CACGAGCGGACATCTGCCATCATACCATCATCATGAAGAAGGCGGACTTAAGGACGAG 334
Qy 82 AsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGly 101
Db 335 AACATCGTTGCTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 394
Qy 102 ValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThr 121
Db 395 GTCATCATTAATCATATCCAGGTCGCGACGCTCTATGCTGGGTGCGCAAGGATTAACAT 454
Qy 122 LysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyVal 141
Db 455 GGGCGAGAGGTCAGAGTCAGCAATTTCTGCGTGTCTGCTGGCAACAACTGCTCTC 514
Qy 142 ThrGlyGlyAsnGlyLysValLysSerGlyProAsnAspAsnIlePheIleTyrTyr 161
Db 515 AGGGTGGGAGCGGCAAGGTTGTGGACAGTGCGCCCAATGACCATATATTTGTTTCTAC 574
Qy 162 AlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLys 181
Db 575 AGTGACATGGGGCTGCGCTGCTGGAATGCTGATGCTATCATCTCTATGCTGAT 634
Qy 182 AspPheAsnGluValLeuGluLysMethHisLysArgLysLysTyrAsnLysMetValIle 201
Db 635 GACCTCTAGATGCTCTGAAGAAGAGCATGCTGCGGGAGCTACAAAAGCGCTGCTTT 694
Qy 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
Db 695 TATCTTGAAGCATGCGAATCTGGAGCATCTTTGAGGCGCTCTCTCCGAATGACATAAT 754
Qy 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTyrGlyValTyrCysProGlu 241
Db 755 GTGATGTCGACCGCGCTCAATCGACAGGAGAGTACTGCGGGAGCTACTGCGCTGGC 814
Qy 242 SerTyrProProSerSerLysGlyThrCysLeuGlyAspThrPheSerIleSer 261
Db 815 GAGTTCGCGGCGCTCCACCGAGTATGACACTGCTTGGGAGACCTGTATAGTGTGCT 874
Qy 262 TrpLeuGluAsnSerAspLeuHisAspMetSerLysGluThrLeuGluGlnIleTyrHis 281
Db 875 TGGATGGAAGACAGTGATTTCCCAATCTGCGAAGTCTGCAAGCAGCAATACAAAC 934
Qy 282 ValValLysArgArgValGlySerAspValProGluThr-----SerHisVal 297
Db 935 TTGGTCAAGGATAGGACAGCG-----GTTGAGGATACATTGAGCTATGCTCCCATGTG 988
Qy 298 CysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsn 317
Db 989 ATGCAATATGTTTCAATGAGTGAATGTTAAGCATCTGTTTTCGTATGTCGACAAAC 1048
Qy 318 ProGluAsnAspAsnPheThrPheThrGluSerPheSerProIleSerAsnSerGly 337
Db 1049 CTGCTTAACGATGACACACGTTTATAGAACAACTCGTTGCCA---TCGTTCTCAAG 1105
Qy 338 LeuValAsnProArgAspIleProLeuGlnArgLysIleGlnLysAlaPro 357
Db 1106 GCTGTTAATCAGCGGACGCTGACCTTCTTCTGCGAGAGTACCGGAAATTCGCA 1165
Qy 358 MetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
Db 1166 GACAGTCACCTGAGAAAAATGAAGCTCGAAGGAGTGTCTGAAGTCATGCGCCACAGG 1225
Qy 378 LysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValGlnThrAsnVal 397
Db 1226 TCTCATGTTGACAGAGTGTGAGCTCATGGAAGCCCTTCTCTTGGCTCTGAGGACGGT 1285
Qy 398 LeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspThrPASPcys 417
```

```
Db 1286 CCAAGGGTTTCTGAAGCCGCTCCGTGCAGCTGTGTGAGCCTCTGTCGATGATTGGAGCTGT 1345
Qy 418 PheLysThrLeuValAsnSerPheLysAsnHisCysGlyValAlaThrValHisTyrGlyLeu 437
Db 1346 CTCAAGTCCACGGTTCGTACTTTTTCAGCGCAATCTGGTCTGTTGGCCCACTATGGGATG 1405
Qy 438 LysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGluThrVal 457
Db 1406 AAGCAGATCGGTCCTTCGCAACATCTGCAACGCTGGCATCTCTCCCTGAAGCAGCTGCG 1465
Qy 458 SerAlaIleGluGlnAlaCysSer 465
Db 1466 AAGTCGCTCTCAGCGCTTCACCC 1489

RESULT 9
US-10-425-114-15845
; Sequence 15845, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15845
; LENGTH: 1769
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-088-P9_FLI
US-10-425-114-15845

Alignment Scores:
Pred. No.: 1 878-112 Length: 1769
Score: 1201.50 Matches: 226
Percent Similarity: 69.16% Conservative: 70
Best Local Similarity: 52.80% Mismatches: 125
Query Match: 48.96% Indels: 7
DB: 8 Gaps: 3

US-09-934-066-2 (1-466) x US-10-425-114-15845 (1-1769)
Qy 42 AlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArg 61
Db 191 GCGTCGGGACCGGTCGTCGTCATCGCGGCTCCACGGCTACTACAACTACCGC 250
Qy 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGlu 81
Db 251 CACGAGCGGACATCTGCCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 310
Qy 82 AsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGly 101
Db 311 AACATCGTTGCTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 370
Qy 102 ValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThr 121
Db 371 GTCATCATTAATCATATCCAGGTCGCGACGCTCTATGCTGGGTGCGCAAGGATTAACAT 430
Qy 122 LysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyVal 141
Db 431 GGGCGAGGCTCAAGCTCGCAACATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490
Qy 142 ThrGlyGlyAsnGlyLysValLysSerGlyProAsnAspAsnIlePheIleTyrTyr 161
Db 491 AGGGTGGGAGCGGCAAGGTTGTGGACAGTGTGCGCCCAATGACCATATATTTGTTTCTAC 550
Qy 162 AlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLys 181
```

```

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15458
; LENGTH: 1774
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-003-Fl2_FLI
US-10-425-114-15458

Alignment Scores:
Pred. No.: 1-88e-112 Length: 1774
Score: 1201.50 Matches: 226
Percent Similarity: 69.16% Conservative: 70
Best Local Similarity: 52.80% Mismatches: 125
Query Match: 48.96% Indels: 7
DB: 8 Gaps: 3

US-09-934-066-2 (1-466) x US-10-425-114-15458 (1-1774)
QY 42 AlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArg 61
Db 196 GCCGTGGGACCGCGTGGCGTGCATCGCGCGCTCCACGCGTACTACAACTACCGC 255
QY 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGlu 81
Db 256 CACCAGCGCGACATCTGCCATGCATACCATGATCATGAAGAAGCGCGACTTAAGGACGAG 315
QY 82 AsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGly 101
Db 316 ACATCGTTGTCTTCTATGATGACATGACATGCGCATAGCGCGGAAATCCGAGCGCTGCT 375
QY 102 ValIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThr 121
Db 376 GTCATATAATCATCCCGAGGTGGCGAGCTATGCTGGGTGCCAAGGATTACACT 435
QY 122 LysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyVal 141
Db 436 GGGCGAGAGGTCAACGTCGACAAATTCCTGCTGCTTGGCAACAAATCGTCTC 495
QY 142 ThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyr 161
Db 496 AGGGTGGGACCGCGCAAGCTTGTGGACAGTGGCCCCCAATGACCATATATTGTTTCTAC 555
QY 162 AlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLys 181
Db 556 AGTGACCATGGGGTCTCGCGTCTTGGAAATGCCTACGTATCATCATCTCTATGCTGAT 615
QY 182 AspPheAsnGluValLeuGluLysMetHisLysArgLysTyrAsnLysMetValIle 201
Db 616 GACCTCGTAGATGCTCTGAAGAAGAACAGCATGCTCGCGGACCTTACAAAAGCTGGTCTTT 675
QY 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
Db 676 TACCTTGAAGCATGCCAATCTGGGACATCTTTGAGGGCTCTCTCCGCAATGACATAAAT 735
QY 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241
Db 736 GTGATGCGACACCGCTCAATGTCAGAGGAGAGTAGTGGGGGACGTACTGCCCTGGC 795
QY 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261
Db 796 GAGTTCGCCGACCGCTCCACCGGAGTAGTACACTTGTCTGGGAGACGTGTATGTTGCT 855
QY 262 TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHis 281
Db 856 TGGATGGAAGACAGTATTTCCCAATCTCGCGAATCTCGCAAGTCAATACAC 915
QY 282 ValValLysArgValGlySerAspValProGluThr-----SerHisVal 297

```

RESULT 10

```

US-10-425-114-15458
; Sequence 15458, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei

```



```

Db 1214 GAAACTGTTGAGCAGAGGAGCATCTTGACAGCAGTATCGATTTCGTGGAGGCTCGTC 1273
Qy 392 ValLysGlnThrAsnValLeuAsnLeuThrSerThrArgThrThrGlyGlnProLeu 411
Db 1274 TTCGGATTGAGAGAGGCGCTTCATGCTGTCAGGCTGTTAGAACCCTCTGGTCAACCATTA 1333
Qy 412 ValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAla 431
Db 1334 GTCGATCAGTGGGACTCTTTGAAGAGGAGTGGTGGGATTTTCGAGTCCGAGTGGGTCA 1393
Qy 432 ThrValHisThrGlyLeuLysThrThrGlyAlaLeuAlaAsnLleCysAsnMetGlyVal 451
Db 1394 CTTACTCAGTACGCGCATGAGTACATGAGAGCGCTTTCGAACACATCTCGAACAGTGGTATA 1453
Qy 452 ---AspValLysGlnThrValSerAlaIleGluAlaCysSer 465
Db 1454 TCTGAGATGAATGAGGAGTCGAGGATC----AGCGCTTGCAGC 1495

```

RESULT 12

```

US-10-425-114-27715
: Sequence 27715, Application US/10425114
: GENERAL INFORMATION:
: APPLICANT: Liu, Jingdong
: APPLICANT: Zhou, Yihua
: APPLICANT: Kovacic, David K.
: APPLICANT: Screen, Steven E.
: APPLICANT: Tabaska, Jack E.
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: CURRENT APPLICATION NUMBER: US/10/425,114
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 73128
: SEQ ID NO 27715
: LENGTH: 1809
: TYPE: DNA
: ORGANISM: zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: LIB4739-001-A9_FLI
US-10-425-114-27715

```

```

Alignment Scores:
Pred. No.: 1.17e-104 Length: 1809
Score: 1125.50 Matches: 225
Percent Similarity: 64.40% Conservative: 68
Best Local Similarity: 49.45% Mismatches: 147
Query Match: 45.86% Indels: 15
DB: 8 Gaps: 5

```

US-09-934-066-2 (1-466) x US-10-425-114-27715 (1-1809)

```

Qy 22 AlaGluSerArgLysThrGlnLeuAsnAspAsnValGluSerSerAspLysSer 41
Db 161 GCGGAGGGTGAGTGGGAGCGCTGATACGAGTCCGACGCGGAGGAGTGCCTGCT 220
Qy 42 Alalys-----GlyThrArgTrpAlaValLeuAlaGly 53
Db 221 TCTGCTCGGCTGCTGAGGATGATGAAGTGGGACGAGGTGGCGGCTGCTGCTGCGCGA 280
Qy 54 SerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspLleCysHisAlaTyrGlnIleLeu 73
Db 281 TCCTTCGTTATGGGAACCTACCGGCACGAGCGCATGTGTGCCATGCGTACCAGATCTG 340
Qy 74 ArgLysGlyGlyLeuLysAspGluAsnIleValPheMetTyrAspAspIleAlaPhe 93
Db 341 CAGAAGGGAGGAGTGAAGAGGAGCAATTCGTGTTGATGATGATGATGATGATGATGATGAT 400
Qy 94 SerSerGluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyr 113
Db 401 AACATTCTGACCCCTAGCCCTGGAGTTATCATTAACCATCTCTAAAGGAGCAATGTTTAT 460

```

```

Qy 114 LysGlyValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnVal 133
Db 461 GATGGTGTTCACAAAGGATTACACAGGTGATCAGTCACTACTGAAAACTTCTTTGCTGTC 520
Qy 134 LeuLeuGlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValLysSerGlyPro 153
Db 521 CTCTTGGGCAATAGAAAGTGTACCACTGAGGAGTAAAGAGGTATAGACAGCAACCT 580
Qy 154 AsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetPro 173
Db 581 AATGACCACATCTTCACTATTACTCAGATCATGGGGTCTCTGGAGTCTTGGTATGCCG 640
Qy 174 ThrGlyAspGluValMetalLysAspPheAsnGluValLeuGluLysMetHisLysArg 193
Db 641 AACCTGCCATATCTGTATGCTGTGCTTCAAGGCTTTGAAAAAGAGACAGTCTTCC 700
Qy 194 LysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGlu 213
Db 701 AATAGCTACTCAAAATGTTATATGTTGAAGCATGTGAAAGTGGCAGTATCTTTGAG 760
Qy 214 GlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSer 233
Db 761 GGGTAAATGCTCAAGATCTAATATTTATGTCAACACAGCCTCAATCCAGTTGAAAT 820
Qy 234 SerTrpGlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCys 253
Db 821 AGTTGGGAACTTACTGCCCTGGGATGGAACCATCACCCCTCTCTGAGTACATTTACCTGT 880
Qy 254 LeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLys 273
Db 881 TTAGTGATTTGACAGTGTCTTCTGGATGGAAGACAGTGAACCCCACTCTTAAGAAG 940
Qy 274 GluThrLeuGlnGlnThrHisValValLysArgArgValGlySerAsp-----Val 291
Db 941 GAAACCATCAAGGATCAGTATGAAGTGTGAAACACAGACCTCAAACTCAATAAGTAC 1000
Qy 292 ProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSer 311
Db 1001 AAAGAGGTTCTCATGTATGAGTATGAGTATGAAGACCTTCAAGGATGAGAGCTTCC 1060
Qy 312 SerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerSer 331
Db 1061 TTTTATCAAGGTTTGTATCTGCTGCTAATGCCAATGTCGAACATGTGCTGTTGGCCAGGC 1120
Qy 332 ProfileSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArg 351
Db 1121 CCA-----AAGGGTGCAGTCAATCAAGAGATGCTGATCTTCTCTCATGTGGAAG 1171
Qy 352 LysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu 371
Db 1172 GGGTACGAGCAGTTAAATGGGGGCGGAGAGAGCTGAGGCGCTCTCATAGAGATTAAA 1231
Qy 372 AspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSer 391
Db 1232 GAAACTGTTCAAGCAGCAAGAGCATCTTGCACAGCATATCGATTTCGTTGGAGGCTCGTC 1291
Qy 392 ValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeu 411
Db 1292 TTCGGATTTCAGAAAGGCGCTTCGATGCTTCGAGGCTTTAGAACCCTCTGGTCAACCATTA 1351
Qy 412 ValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAla 431
Db 1352 GTCGATGACTGGGACTCTTTGAAGAGGAGTGGTGGGATTTTCGAGTCCAGTGTGGTTCA 1411
Qy 432 ThrValHisThrGlyLeuLysTyrThrGlyAlaLeuAlaAsnLleCysAsnMetGlyVal 451
Db 1412 CTTACTCAGTACGCGCATGAGTACATGAGAGCGCTTTCGAACACATCTCGAACAGTGGTATA 1471
Qy 452 ---AspValLysGlnThrValSerAlaIleGluAlaCysSer 465
Db 1472 TCTGAGATGAATGAGGAGTCGAGGATC---AGCGCTTGCAGC 1513

```

RESULT 13


```

Db      1375 ATTCCGGAGACAGTGAACCATAGAACACATAGATCGTAGCGGTGGAATTGCATTGGAGTT 1434
Qy      390 LeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrGlycIn 409
          |||               :|||   |||   |||   |||   |||   |||
Db      1435 TTACTGTATGCACCAGAAAAGGTCTTCCTGCTTCAACAATCCGTGAGGGCTCCTGGTTTCG 1494
Qy      410 ProLeuValAspAspTrpAspCysPheLysThrLeuValAlaAsnSerPheLysAsnHisCys 429
          |||||||         |||   |||   |||   |||   |||   |||
Db      1495 TCCCTCTGTTGATGACTGGACATGCCTAAATAAATCAATGGTTCGGGTGTTTTGAAACACTCAGT 1554
Qy      430 GlyAlaThrValHisTyrgLyLeuLysTyrrThrGlyAlaLeuAlaAsnIleCysAsnMet 449
          |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db      1555 GGGACACTGACTCAGTATGGCATGGAACACATCGCAGCATTCGCCACATTTTGCAACAGT 1614
Qy      450 GlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCys 464
          |||||       :|||   |||   |||   |||   |||   |||
Db      1615 GGCCTT-----TCTGAGGCCTCCATGGAAGAGGCTTCT 1647

RESULT 15
US-10-425-114-10538
; Sequence 10538, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 10538
; LENGTH: 1651
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700942962_FU1
US-10-425-114-10538

Alignment Scores:
Pred.: NO.:                2.58e-98           Length:        1651
Score:                    1063.00             Matches:         200
Percent Similarity:      68.24%              Conservative:     60
Best Local Similarity:   52.49%              Mismatches:    119
Query Match:            43.32%               Indels:          2
DB:                      8                   Gaps:           2

US-09-934-066-2 (1-466) x US-10-425-114-10538 (1-1651)
Qy      85 ValPheMetTyraSpAspIleAlaPheSerSerGluAsnProArgProGlyValIlelle 104
          |||||||         |||   |||   |||   |||   |||   |||
Db      1 GTTTTATGATGATGATTCATTCATTAATGGGAAAAACCAAGCCCTGGAGTCATCAT 60
Qy      105 AsnLysProaspGlyGluAspValTyrrLysGlyValProLysAspTyrrThrLysGluAla 124
          |||||||         |||   |||   |||   |||   |||   |||
Db      61 AACAAACCATGAGGTGATGTTTATAAGAGATTCCAAAGGATTACACCCGGCGAAGAT 120
Qy      125 ValAsnValGlnAsnPheTyrrAsnValLeuLeuGlyAsnGluSerGlyValThrGlyGly 144
          |||   |||   |||   |||   |||   |||   |||   |||
Db      121 GTTACTGTTGATAACTTTTTTGGCTGCTTTACTTGGAAATAAGTCAGCACACTGACTGGTGGC 180
Qy      145 AsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrrAlaAspHis 164
          :|||   |||   |||   |||   |||   |||   |||   |||
Db      181 AGTGGGAAGGTGTGGACAGTGGCTCATGATCATATATTTGTATATACTACTACCAT 240
Qy      165 GlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsn 184
          |||   |||   |||   |||   |||   |||   |||   |||
Db      241 GGAGGTCTGGGGTGTGGCGGATGCCCTGGCTCCTACTTACTATACGGCGGATGATCTGATT 300
Qy      185 GluValLeuGluLysMetHisLvsArgLysLvsTyrrAsnLysMetValIleTyrrValGlu 204

```

```
Db 301 GAACTCTTGAAGAAAAGCATCTTCTGGAACATATAAAACCTAGTATTTTATCTCTGCAG 360
Qy 205 AlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAla 224
Db 361 GCATGTGAATCTGGAGTATCTTTGAAGCTCTTCCTGAAGATATCAATATTTATGCA 420
Qy 225 ValThrAlaAlaAsnSerLysGluSerSerTyrGlyValTyrCysProGluSerTyrPro 244
Db 421 ACCACTGCTTCCAAATGCAGAAGAAATAGTTGGGACATATTTGCCCGGGAGTATCCT 480
Qy 245 ProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGlu 264
Db 481 AGCTCTCCCCCAAGATATACAACTCTTTGGGTGAGCTTGTACAGTGTCTCTGGATGGAA 540
Qy 265 AspSerAspLeuHisMetSerLysGluThrLeuGluGlnGlnTyrHisValValLys 284
Db 541 GACAGTGACAGACAAATTTGGCAACAGAACTCTGCACCAACAATATAAATTTGTTAAA 600
Qy 285 ArgArg---ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGlu 303
Db 601 GAGAGGACTATATCTGGAGATTCTACTATGCTCTCACGTGATGCATGTGGTGATGTA 660
Qy 304 LysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPhe 323
Db 661 GGGCTTAGCAGAGATGTTCTTCCATTATTTGGGTACAGATCTCTGCTAATGATAATTC 720
Qy 324 ThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProArgAsp 343
Db 721 ACTTTTGTGGATGAA---AACTCCTTATGCTCACCTTCAAAACACAGTCAACCAACGTCAT 777
Qy 344 IleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSer 363
Db 778 GCTGATCTCATCCATTTTGGGATAAGTCCGCAAAAGCTCCTGAGGTTCTCTCAGGAA 837
Qy 364 LysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSer 383
Db 838 AATACAGCTCAGAAACAAGTTTGGGAAGCAATCTCTCAGAGATGCATGATAGACACAGT 897
Qy 384 IleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSer 403
Db 898 GTAAACTGATTTGGGAGCTTTTATTTGGCATTGAAAGGGTCCAGAAAGTACTCAACGCT 957
Qy 404 ThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThrLeuValAsn 423
Db 958 GTTAGACCGGCTGGATCGGCACCTGTGTGAGTACTGGCCTGAAACCAATGTTGAGG 1017
Qy 424 SerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeu 443
Db 1018 ACTTTTGAGACACATTGTGGATCTTGTCTCAATACGGGATGAACACATGAGGTCCTTT 1077
Qy 444 AlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGlnAla 463
Db 1078 GCAACATCTGCAATGTAGGATAAAGAAATGAACAATGGCTGAGGCTTCAGCACAAGCT 1137
Qy 464 Cys 464
Db 1138 TGT 1140
```

Search completed: May 29, 2003, 20:40:32
Job time : 719 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 18:04:19 ; Search time 1602 Seconds
(without alignments)
4711.047 Million cell updates/sec

Title: US-09-934-066-2
Perfect score: 2454
Sequence: 1 MSSPLGHFQILVFLHALLIF.....CNMGVDVKQTVSAIEQACSM 466

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US0934066/runat_19052003_163700_24361/app_query.fasta_1.647
-DB=EST -QPM=fastap -SUFFIX=p2n.est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biolum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pcpt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0934066.ecgn_1.1.2874.erunat_19052003_163700_24361 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmd:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	1204.5	49.1	1981	11	AV104034	AV104034 Zea mays	
2	1089.5	44.4	1651	11	AV110063	AV110063 Zea mays	
3	968	39.4	541	10	AV559545	AV559545 AV559545	
C	4	924	37.7	606	10	AV557684	AV557684
5	922	37.6	760	14	BQ871979	BQ871979 OGI113R11	
6	921	37.5	746	14	BQ871135	BQ871135 OGI10P16	
7	907	37.0	796	12	BG645154	BG645154 EST50677	
8	904	36.8	711	12	BG646037	BG646037 EST50765	
9	903	36.8	666	12	BG411524	GA__Ea00	
10	896	36.5	766	12	BG123960	BG123960 EST46960	
C	11	893	36.4	585	10	AV559121	AV559121
12	869	35.4	776	12	BG581593	EST48332	
13	867	35.3	561	10	AV559478	BG581593	
C	14	855	34.8	538	10	AV555212	AV559478
15	848	34.6	535	10	AV562667	AV555212	
16	818	33.3	669	10	AV940380	AV562667	
17	817	33.3	574	10	AV831492	AV940380	
18	815.5	33.2	801	13	BI308524	AV831492	
19	814	33.2	655	14	BQ871920	BI308524	
20	804	32.8	733	14	BU028909	BQ871920	
21	802	32.7	679	10	AV926752	BU028909	
C	22	784	31.9	506	10	AV557191	AV926752
23	782	31.9	719	14	BQ841709	AV557191	
24	758	30.9	431	10	AV564659	BQ841709	
25	752	30.6	617	14	BQ579190	AV564659	
26	752	30.6	718	12	BG129376	BQ579190	
27	749	30.5	845	12	BF276350	BG129376	
28	745	30.4	836	12	BF263324	BF276350	
C	29	743	30.3	535	10	AV561906	BF263324
30	742	30.2	646	14	BU025783	AV561906	
31	739	30.1	638	10	BE660642	BU025783	
32	738	30.1	654	10	AW224404	BE660642	
33	732.5	29.8	1345	11	AY106221	AW224404	
34	730.5	29.8	716	10	BE660892	AY106221	
35	730	29.7	919	14	BQ881734	BE660892	
36	725	29.5	768	14	BU004552	BQ881734	
37	722.5	29.4	1057	9	AL544478	BU004552	
38	716.5	29.2	681	10	AW224405	AL544478	
39	716	29.2	661	14	BQ148226	AW224405	
40	714.5	29.1	765	12	BG591158	BQ148226	
41	704	28.7	595	9	AJ436163	BG591158	
42	700.5	28.5	861	14	BQ936258	AJ436163	
43	696.5	28.4	755	14	BU002036	BQ936258	
44	696	28.4	865	13	BI953500	BU002036	
45	695	28.3	600	14	BQ997181	BI953500	
						HVSMem001	
						OQGI14P12	

ALIGNMENTS

RESULT 1
AY104034
LOCUS
DEFINITION Zea mays PC0131778 mRNA sequence.
ACCESSION AY104034
VERSION AY104034.1 GI:21207112
KEYWORDS HPC.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1981)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 JOURNAL Unpublished (2002)
 REFERENCE 2 (bases 1 to 1981)
 AUTHORS Coe, E.C.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 FEATURES Location/Qualifiers
 source 1..1981
 /organism="Zea mays"
 /db_xref="taxon:637688"
 /clone_lib="PCO131778"
 /clone_lib="Maize Mapping Project/DuPont Consensus
 Library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

BASE COUNT 480 a 509 c 505 g 487 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4,28e-137 Length: 1981
 Score: 1204.50 Matches: 227
 Percent Similarity: 69.16% Conservative: 69
 Best Local Similarity: 53.04% Mismatches: 125
 Query Match: 49.08% Indels: 7
 DB: 11 Gaps: 3

US-09-934-066-2 (1-466) x AY104034 (1-1981)

Qy 42 AlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArg 61
 Db 203 GCCGTGGGACCGGTCGCGGTCATCGCGGCTCCAGCGCTACTACAACTACCCG 262
 Qy 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGlu 81
 Db 263 CACGACGGGACATCGCCATGATACACAGATCATGAAGAAGGGCGGACTTAAGACGAG 322
 Qy 82 AsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGly 101
 Db 323 AACATCGTGTCTTCATGTACGATGACATCGCGCATAGCCCGGAAATCCGAGGCGCTGT 382
 Qy 102 ValIleLeuAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThr 121
 Db 383 GTCATCAATCAATCCCGAGGTCGCGGACGCTATGCTGGGGTGCCCAAGGATTACACT 442
 Qy 122 LysGluAlaValAsnValGlnAsnPhenTyrAsnValLeuLeuGlyAsnGluSerGlyVal 141
 Db 443 GGGCGAGAGGTCAACGTCGACATTTCTTCGCTGTCTGTTGGCAACAAACTGCTCTC 502
 Qy 142 ThrGlyGlyAsnGlyLysValLysSerGlyProAsnAspAsnIlePheIleTyrTyr 161
 Db 503 AGGGGTGGGAGCGGCAAGCTTGTGGACAGCTGCGCCCAATGACCATATATTTGTTTCTAC 562
 Qy 162 AlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLys 181
 Db 563 AGTGACCATGGGGTCTCGGCTCTTGAATGCCATACGTATCATATCTCATGCTGAT 622
 Qy 182 AspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValle 201
 Db 623 GACCTCGTATGCTCTGAAGAAGAGCATGTCGCGGACCTACAAAGCGCTGCTT 682
 Qy 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
 Db 683 TATCTTGAAGCATGCCAAATCTGGGAGCATCTTTGAGGCGCTCTCCCGAATGACATAAT 742
 Qy 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241

Db 743 GTGTATCGGACCACCGCTCAATGCAGAGGAGTAGCTGGGGAGCTACTTCCCTGCG 802
 Qy 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261
 Db 803 GAGTTCGCGACCGCTCCCGGAGTAGACATCTGCTGGGAGACCTGTATAGTGTCT 862
 Qy 262 TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHis 281
 Db 863 TGGATGGAGACAGATGATTTCCAAATCTCGCAACTGAATCTCTCAAGCAGCAATACAAC 922
 Qy 282 ValValLysArgArgValGlySerAspValProGluThr-----SerHisVal 297
 Db 923 TTGGTCAAGATAGACAGCG-----GTTCAAGATACATTCAGCTATGCTCCCATGTG 976
 Qy 298 CysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsn 317
 Db 977 ATGCAATATGCTTCATTTGGAGTTCAATGTAAAGCATCTGTTTCGPACATTCGCCACAAAC 1036
 Qy 318 ProGluAsnAspAsnPheThrGluSerPheSerSerProIleSerAsnSerGly 337
 Db 1037 CCTGCTAACCATGACAAACACGTTTATAGAAGACAACTCGTTGCCA---TCGTTCTCAAAAC 1093
 Qy 338 LeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaPro 357
 Db 1094 GCTGTTAATCAGCGCGACGCTGCTTCTACTTCTGGCAGAGTACCGGAATTTGGCA 1153
 Qy 358 MetCysSerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
 Db 1154 GACAGCTACCTGAGAAAATGAAGCTCGAAGAGTTCCTTGAAGTATGCGCCACAGG 1213
 Qy 378 LysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnVal 397
 Db 1214 TCTCATGTTTACACAGCATGTTGAGCTCATTTGGAAGCTTCTCTTTGGCTCTGAGGACGGT 1273
 Qy 398 LeuAsnLeuLeuThrSerThrArgThrGlnProLeuValAspAspTrpAspCys 417
 Db 1274 CCAAGGTTCTGAAGACCGCTCCGTCGAGCTGGTGAGCGCTCTGCTGATGATTGGAGCTGT 1333
 Qy 418 PheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeu 437
 Db 1334 CTCAAGTCCACAGGTCGTACTTTGAGGCGCAATGTGGTCTGTTGGCGCAGTATGGGATG 1393
 Qy 438 LysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrVal 457
 Db 1394 AAGCATGCGGTCCTTCGCAACATCTCGAAGCTGGCATCTCTTCTCAAGCAGTGTCTG 1453
 Qy 458 SerAlaIleGluGlnAlaCysSer 465
 Db 1454 AAGTCGCTGCTCAGGCTTGACCC 1477
 RESULT 2
 AY110063
 LOCUS
 DEFINITION Zea mays CL863_1 mRNA 'sequence'.
 ACCESSION AY110063
 VERSION AY110063.1 GI:21214138
 KEYWORDS HTG.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 Clade; Panicoideae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 1651)
 AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
 Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
 TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 JOURNAL Unpublished (2002)
 REFERENCE 2 (bases 1 to 1651)
 AUTHORS Coe, E.C.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA

/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 177 a 88 c 128 g 148 t
ORIGIN

Alignment Scores:
Pred. No.: 9,99e-109 Length: 541
Score: 968.00 Matches: 180
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39,45% Indels: 0
DB: 10 Gaps: 0

US-09-934-066-2 (1-466) x AV559545 (1-541)

QY 96 GluAsnProGluValIleLeuAsnLysProAspGlyGluAspValTyrLysGly 115
DB 2 GAGATCCCTAGCCCTGGAGTTATCATTAATACCCAGATGGAGAAGATGTTTATAAGGA 61
QY 116 ValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeu 135
DB 62 GTTCCTAGGACTACACTAAGAGCTGTTAATGTCAAACTTCTACAATGTGTACTT 121
QY 136 GlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAsp 155
DB 122 GGAATGAAGTGGCTCAGAGGAAATGCAAAAGTGTGCAAAAGTGGTCCCTAATGAT 181
QY 156 AsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
DB 182 ATATCTTCATATATGTCTGACCATGGAGCTCTGCTTAAATAGCGATCCCACTGGT 241
QY 176 AspGluValMetAlaLysAspPheAsnGluValLeuGlyLysMetHisLysArgLysLys 195
DB 242 GATGAAGTTATGCAAAAGATTTCATGAAGTCTTGGAGAAGATGCATTAAGAGAAAAA 301
QY 196 TyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyTle 215
DB 302 TACAACAAGATGGTATCTATCTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGGATT 361
QY 216 LeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTyr 235
DB 362 TTAAGAAAAATCTCAACATATACGAGTGAAGTCTGCTAATCTTAAAGAGAGCAGCTG 421
QY 236 GlyValTyrCysProGluSerTyrProProSerGluIleGlyThrCysLeuGly 255
DB 422 GGAGTTTACTGTCTGAGTCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGC 481
QY 256 AspThrPheSerIleSerTyrLeuGluAspSerAspLeuHisAspMetSerLysGluThr 275
DB 482 GATACATTAGCATCTCTGGCTTCAGGACAGTGCACCTTCATGACATGACGCAAGAGACT 541

RESULT 4
AV557684/c 606 bp mRNA linear EST 07-SEP-2000
LOCUS
DEFINITION
Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone SQ075c12f 3', mRNA sequence.

ACCESSION
AV557684
VERSION
AV557684.1 GI:8729099

KEYWORDS
EST.

SOURCE
thale cress..

ORGANISM
Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 606)

Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries

DNA Res. 7, 175-180 (2000)

JOURNAL
MEDLINE
20363093

COMMENT

Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source
1..606

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ075c12f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 158 a 110 c 124 g 214 t
ORIGIN

Alignment Scores:
Pred. No.: 3.18e-103 Length: 606
Score: 924.00 Matches: 177
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.65% Indels: 0
DB: 10 Gaps: 0

US-09-934-066-2 (1-466) x AV557684 (1-606)

QY 290 AspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyr 309
DB 606 GATGTACCAAGAGACTTCTCATGTATGCCCTTTCCGAACAGAGAAGATGCTTAAAGATTAT 547
QY 310 LeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPhe 329
DB 546 CTTTCTCTTACATTTCGAAGAATCTGANAACGATACTTCACTTTTCCGGAATCCTTT 487
QY 330 SerSerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeu 349
DB 486 TCCTCACCACATCTCTAATTCCTGGCTTGGTCAATCCGCGGATATCTCTCTGTATACCTC 427
QY 350 GlnArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLys 369
DB 426 CAGAGAAGATTCAAAAAGCTCCAAATGGCATCCTTGAAGCAAGAGTCCAGAAGAAA 367
QY 370 LeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArg 389
DB 366 TTGCTTGACGAAAAGATCATAGGAACAAATCGATCAGAGCATTTACAGACATTTCTCGCG 307
QY 390 LeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrGlyGln 409
DB 306 CTTTCAGTTAAACAAACCAATGCTTAAATCTCTTAACCTCCACAAAGAACACACAGACAG 247
QY 410 ProLeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCys 429
DB 246 CCTCTGTACAGCATGGGATTCCTCAAGACTCTAGTTAATAGCTTCAAGAAATCAGTCC 187
QY 430 GlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMet 449
DB 186 GGTGCAACGGTGCATTTACGGATTGAAGTATACAGAGCGCTTGCCTCAATATCTGCANTATG 127
QY 450 GlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSerMet 466
DB 126 GGAGTGGATGTGAAGCAAACTCTTTTACGCCATTTGAACAGCTTGTTCGATG 76

RESULT 5

BQ871979

LOCUS

DEFINITION

BQ871979

ACCESSION

VERSION

BQ871979.1

KEYWORDS

SOURCE

BQ871979 760 bp mRNA linear EST 15-AUG-2002
OG113F11.yg.abl OG_ABCDI lettuce salinas Lactuca sativa cDNA clone
OG113F11, mRNA sequence.
BQ871979
BQ871979.1 GI:22258529
EST.
Lactuca sativa.

```

ORGANISM  Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.

REFERENCE 1 (bases 1 to 760)
AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
JOURNAL   Unpublished (2002)
COMMENT   Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig7059, see http://cgpdb.ucdavis.edu/
for details.
Plate: QG113 row: F column: 11.

FEATURES             Location/Qualifiers
     source            1..760
                        /organism="Lactuca sativa"
                        /cultivar="Salinas"
                        /db_xref="taxon:4236"
                        /clone="QG113f11"
                        /clone_lib="QG_ARCDI lettuce salinas"
                        /lab_host="E.coli"
                        /note="Vector: pBRCNAsfiAB; The library was constructed
                        from 10 different sources of RNA from a single genotype.
                        Separate cDNAs were generated using primers that
                        incorporated unique 5' and 3' tags to distinguish each
                        source of RNA. cDNAs were then pooled, size-fractionated,
                        directionally cloned into a custom medium-copy vector and
                        transformations made with four size classes to minimize
                        size bias. Details of each source of RNA and library
                        construction can be obtained at http://cgpdb.ucdavis.edu/
                        TAG_LIB-QG_ARCDI lettuce salinas
                        TAG_TISSUE=roots
                        TAG_SEQ=GTTCACGGG"

BASE COUNT  232 a 131 c 190 g 207 t
ORIGIN
Alignment Scores:
Pred. No.:      7.8e-103      Length:      760
Score:          922.00      Matches:    162
Percent Similarity: 78.86%      Conservative: 32
Best Local Similarity: 65.85%      Mismatches: 52
Query Match:    37.57%      Indels:     0
DB:             14      Gaps:      0

US-09-934-066-2 (1-466) x BQ871979 (1-760)
QY 44 GlyThrargTrpAlaValLeuAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
Db 15 GGTACACGATGGGCTGATTGCTGCTGGATCTAATGGCTACTGGAATTACACACATCAG 74
QY 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
Db 75 GCTGATGATGCGCATGCATACCAAACTTGAAGAAAGGTGGCATTAAGATGAAGAAACATT 134
QY 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
Db 135 GTGGTTTTCATGATGATGACATGTCAAATTAATGAGGAGAACCCCGCCCTGGAGTCATT 194
QY 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrTyrLysGlu 123
Db 195 ATTAATAGCCAGATGGTGATGATGATATCATGGAGTCCCAAGAGGATTACATGGGGAA 254
QY 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143

```

```

Db 255 GATGTCACCTGGACACACTTCTTTGCTGTTCTGGAATAAAAGTCAGTTAAAGGG 314
QY 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
Db 315 GGAAGTGAAGAGTGTGGACGAGTGGCCCAATGATCATCATCTTTATATATATACCGAT 374
QY 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
Db 375 CATGGTGTCCAGGTGTGCTCGGGATGCCAACAAATCCTTACATGTATGTCTAATGATCTG 434
QY 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 435 ATTGAGGTCTTGAAGAAAAACATGCTTCTGGAACATACAAAGCATGGTATTATCTT 494
QY 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
Db 495 GAAGCATGTAATCTGGAAGTATATTCGAGGAGCTTCTCCCTGAGGGTTTAAATATATAT 554
QY 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
Db 555 GCAACGACAGCATCTGCTCCAGATGAGAACAGTTGGGGTACTTACTGCCCTGGAGATTAT 614
QY 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
Db 615 CCCAGTCCACCTCCAGAAATATGATACCTGTTGGGTGATTGTACAGCTGTTCATGATG 674
QY 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal 283
Db 675 GAGACTGTGATGTACACAACTCGCGAACAAACATCAGGACGACCAATATCATCTCGTT 734
QY 284 LysArgArgValGlySer 289
Db 735 AAAGAAAGAACCTCAAT 752

RESULT 6
BQ871135 746 bp mRNA linear EST 15-AUG-2002
LOCUS   QG110P16.yg.abl QG_ARCDI lettuce salinas Lactuca sativa cDNA clone
DEFINITION   QG110P16, mRNA sequence.
ACCESSION   BQ871135
VERSION     BQ871135.1 GI:22257685
KEYWORDS    EST.
SOURCE      Lactuca sativa.
ORGANISM    Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE   1 (bases 1 to 746)
AUTHORS     Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
TITLE       Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
JOURNAL     Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig7059, see http://cgpdb.ucdavis.edu/
for details.
Plate: QG110 row: P column: 16.
source      Location/Qualifiers
            1..746
            /organism="Lactuca sativa"
            /cultivar="Salinas"
            /db_xref="taxon:4236"
            /clone="QG110P16"

```

/clone_lib="QG-ABCDI lettuce salinas"
/lab_host="E.coli"

/note="Vector: pBRCDNASFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transfections made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>

TAG_LIB=QG-ABCDI lettuce salinas
TAG_TISSUE=ROOTS
TAG_SEQ=GTTCACGGG"

BASE COUNT 226 a 129 c 187 g 204 t
ORIGIN

Alignment Scores:
Pred. No.: 1,018-102 Length: 746
Score: 921.00 Matches: 162
Percent Similarity: 79.42% Conservative: 31
Best Local Similarity: 66.67% Mismatches: 50
Query Match: 37.53% Indels: 0
DB: 14 Gaps: 0

US-09-934-066-2 (1-466) x B0871135 (1-746)

Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
Db 15 GGTACACGATGGCTGTATTGCTGCTGATCTAATGCTACTGGAATATACAGACATCAG 74
Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIle 83
Db 75 GCTGATGATGCATGCATACCAATCTTGAAAAAGCTGGCATTAAAGGATCAAAACATT 134
Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
Db 135 GTGGTTTCATGATGATGACATTCCTCAATAATAGGAGAACCCAGCCCTGGATCATT 194
Qy 104 IleAsnLysProAspGlyCysValTyrLysGlyValProLysAspTyrThrLysGlu 123
Db 195 ATTATAGCCAGATGGTGTATGATGATCATGATGATGATGATGATGATGATGATGATG 254
Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
Db 255 GATGACTGTGGACACTTCTTCTGCTGCTGGAATAAAGTCAGCTTAAGGG 314
Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
Db 315 GGAAGTGGAAAGGTGTGAGCAGTGGCCCAATGATCATCTTTATATATATATACCGAT 374
Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
Db 375 CATGCTGTCCAGGTGTGCTGGGATGCCAAATCCTTACATGCTATGCTAATGATCTG 434
Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 435 ATTGAGGTCTTGAAGAAAAACATCTCTGCAACATACAAAGCATGGTATTTTATCTT 494
Qy 204 GluAlaCysGlySerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
Db 495 GAAGCATGTGAATCTGGAATATATTCAGGAGCTCTCTCCCTGAGGTTTAAATATATAT 554
Qy 224 AlaValThrAlaAlaAsnSerLysCysSerSerPheGlyValTyrCysProGluSerTyr 243
Db 555 GCAACGACAGCATCTGCTCCAGATGAGAACAGTGGGGTACTTACTGCCCTGGAGATAT 614
Qy 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
Db 615 CCCAGTCCACCTCCAGAAATATGATACCTGTGTGGTGAATTTGATGATGTTGCATGATG 674
Qy 264 GluAspSerAspLeuHisAspMetSerLysGluThrLysGluGlnGlnTyrHisValVal 283

Db 675 GAGGACTGTGATGTACACAACTCGACAGAAACAAATCAGGACGAATATCATCTGGTT 734
Qy 284 LysArgArg 286
Db 735 AAAGAAAGA 743
RESULT 7
LOCUS BG645154 796 bp mRNA linear EST 24-APR-2001
DEFINITION EST506773 KV3 Medicago truncatula cDNA clone pKV3-39K10 5' end,
mRNA sequence.
ACCESSION BG645154
VERSION BG645154.1 GI:13780266
KEYWORDS EST,
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;
Medicago.
REFERENCE 1 (bases 1 to 796)
AUTHORS VandenBosch,K., Endre,G., Hur,J., Beremand,P., Town,C.D., Bowman
C.L., Craven,M.B., Cho,J. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula 72 h after Rhizobium
inoculation, 2001
JOURNAL Unpublished (2001)
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
M393561e TIGR sequence name: MTEBX65TK More information is
available at: www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
Location/Qualifiers
1..796
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pKV3-39K10"
/tissue_type="Seedling roots"
/dev_stage="3 days post-inoculation with Sinorhizobium
meliloti"
/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XLOLR cells."

BASE COUNT 229 a 136 c 193 g 238 t
ORIGIN

Alignment Scores:
Pred. No.: 5,92e-101 Length: 796
Score: 907.00 Matches: 161
Percent Similarity: 79.60% Conservative: 38
Best Local Similarity: 64.40% Mismatches: 50
Query Match: 36.96% Indels: 1
DB: 12 Gaps: 0

US-09-934-066-2 (1-466) x BG645154 (1-796)

Qy 38 SerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyr 57
Db 27 AATGATGACAAACGATCAAGGACTAGGTGGCTATTTACTTGGTGTCTAATGTTAT 86
Qy 58 TyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGly 77

taken from greenhouse plants (4-6wks old TA496). Tissue was immediately frozen in liquid nitrogen."

```

BASE COUNT      216 a 136 c 186 g 228 t
ORIGIN

Alignment Scores:
Pred. No.:      1,28e-99      Length:      766
Score:          896.00        Matches:     162
Percent Similarity: 79.05%    Conservative: 38
Best Local Similarity: 64.03% Mismatches:    52
Query Match:     36.51%      Indels:       2
DB:              12          Caps:         1

US-09-934-066-2 (1-466) x BGL123960 (1-766)

QY 50 LeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAla 69
DB 2 CTTCTGCCGGATCAATGTTATTTGAATATAGACATCAGGCTGATGTATGCCATGCG 61
QY 70 TyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMetTyrAsp 89
DB 62 TATCAGCTATTGAGAAAGGTGCTCAAGATGAAATATTATTGCTTCATGTATGAT 121
QY 90 AspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAspGly 109
DB 122 GACATTGCTCACCATCAAGAGAACCCCAAGACCAGAGATTATTATTATAGTCTCGCGGT 181
QY 110 GluAspValTyrLysGlyValProLysAspTyrTyrLysGluAlaValAsnValGlnAsn 129
DB 182 GAGATGTTTACGAAGGATCTCTAGGATTACCGGGAGATGATGTTAATGTGCACAC 241
QY 130 PheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValVal 149
DB 242 TTTTGTAGCTGCTCTCTGTTGTAACAACAGCTCTTACTGAGGTAGCGGAAGGTGGTG 301
QY 150 LysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeu 169
DB 302 AATAGTGGTCCCAATGATCATATTTTCATATTTATAGTATGATGATGTCGCCCTGCGGT 361
QY 170 IleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuGluLys 189
DB 362 CTGGGATGCTACCAATCTTATCTATATGCGGATGATCTAATGCTGTGTTGAAAAG 421
QY 190 MethisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGly 209
DB 422 AAGCATGCCCTGGGACATATAAAGCTTGATTTGTTACATTTGAAGCTTCGAGTCTGGA 481
QY 210 SerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsn 229
DB 482 AGTATATTTGAGGACTCTCTCTTAATGCTTAATATTTATGCCACACAGCTTCAAT 541
QY 230 SerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProSerGlu 249
DB 542 GCTGAGAAGACGAGCTGGGAACCTATTGCTCGGAGATATCTAGTCTCTCTCTCGAA 601
QY 250 IleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHis 269
DB 602 TATCAGACTTGTGCTGGGTATTTGATGCTGCTTCTGGATGAGACAGTGAATGCAC 661
QY 270 AspMetSerLysThrLeuGluGlnGlnTyrHisValValLysArgArgValGlySer 289
DB 662 AACTTCGGGACTGAAATTTTGGAGCAGCAGTATCCTTGTGTCAAAAGAGAACTGCAATG 721
QY 290 AspVal---ProGluThrSerHisValCysArgPheGly 301
DB 722 GAATACTGCCTAT-GGTTCCCATGTCATGCAATTTGGT 759

RESULT 11
AV559121/c 585 bp mRNA linear EST 07-SEP-2000
LOCUS
DEFINITION AV559121 Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQU11e10F 3', mRNA sequence.
ACCESSION AV559121

```

```

VERSION AV559121.1 GI:8730547
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 585)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
       of 12,028 non-redundant expressed sequence tags from normalized and
       size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
       The First Laboratory for Plant Gene Research
       Kazusa DNA Research Institute
       Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
       Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
       Location/Qualifiers
FEATURES             1..585
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SQU11e10F"
                     /clone_lib="Arabidopsis thaliana green siliques Columbia"
                     /tissue_type="green siliques"
                     /notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT 160 a 110 c 118 g 197 t
ORIGIN

Alignment Scores:
Pred. No.:      2,02e-99      Length:      585
Score:          893.00        Matches:     171
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:     36.39%      Indels:       0
DB:              10          Caps:         0

US-09-934-066-2 (1-466) x AV559121 (1-585)

QY 296 HisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGly 315
DB 585 CATGTATGCCGTTTCGGAACAGAGAAGATGCTTAAGATTATCTTCTCTTACATTGGA 526
QY 316 ArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerProIleSerAsn 335
DB 525 AGAATCTCTGAAACGATACTTCACTTTTCACGGAATCTTTTCTCCCAATCTCTAAT 466
QY 336 SerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLys 355
DB 465 TCTGGCTTGTCTCAATCCGCCGATATTCCTCTGCTATACCTCCAGAGAAGATTCAAAA 406
QY 356 AlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsn 375
DB 405 GCTCCAATGGCATCACTTGAAGCAAAAGAGCTCAGAGAAGAAATGCTTGACGAAACAAT 346
QY 376 HisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThr 395
DB 345 CATAGGAACAAATTCGATCAGACATTTACAGACATTTCTGCGGCTTTTCAGTTAAACAACC 286
QY 396 AsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrp 415
DB 285 AATGCTCTTAATCTCTTAACCTCCACAAGAACACAGGACAGCTCTTTGTAGACGATTGG 226
QY 416 AspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyr 435
DB 225 GATTGCTTCAAGACTCTAGTTAATAGTTCACGAATCACTGCGGTGCAACGGTCAATTAC 166
QY 436 GlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGln 455
DB 165 GGATTGAAGTATACAGGAGCGCTTGCCAATATCTGCAATATGAGAGTGGATGTGAACGAA 106

```



```

/clone="S0118b02F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 154 a 105 c 113 g 189 t
ORIGIN
Alignment Scores:
Pred. No.: 3,08e-96 Length: 561
Score: 867.00 Matches: 167
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.33% Indels: 0
DB: 10 Gaps: 0
US-09-934-066-2 (1-466) x AV559478 (1-561)
QY 300 PheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGlu 319
Db 559 TTCGGAACAGAGAGATGCTTAAGATTATCTTCTCTTACATTGCAAGAAATCTGAA 500
QY 320 AsnAspAsnPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuVal 339
Db 499 AACGATAACTTCACTTTCACGAATCTTCTCCTACCAATCTCTAATCTGGCTTGGTC 440
QY 340 AsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGly 359
Db 439 AATCCGCGGATATCTCTGCTATACCTCCAGAGAAAGATTCAAAAGCTCCAAATGGA 380
QY 360 SerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGln 379
Db 379 TCACCTTGAAGCAAGAGAGCTCAGAGAAATGCTTGAGCAAGAAATCATAGGAAACAA 320
QY 380 IleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsn 399
Db 319 ATCGATCAGAGCATACACACATCTCGCGCTTTCAGTTAAACAAACCAATGCTTAAAT 260
QY 400 LeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLys 419
Db 259 CTCCTTAATCCACAAGAACACAGACAGACAGCTCTTGTAGACGATTGGGATTCCTCAAG 200
QY 420 ThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyr 439
Db 199 ACTCTAGTTAATAGCTTCAAGAAATCTCGGTGCAACGGTGCAATACGAGATTGAAGTAT 140
QY 440 ThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAla 459
Db 139 ACAGGAGCGCTGCGCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTCAGCC 80
QY 460 IleGluGlnAlaCysSerMet 466
Db 79 ATTGAACAAGCTGTGTCGATG 59
RESULT 14
AV555212 538 bp mRNA linear EST 06-SEP-2000
LOCUS Arabidopsis thaliana green siliques Columbia Arabidopsis
DEFINITION thaliana cDNA clone S0008c10f 3', mRNA sequence.
ACCESSION AV555212
VERSION AV555212.1 GI:8726627
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 538)
Asanizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
JOURNAL
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
Location/Qualifiers
1..538
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="S0008c10f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 168 a 96 c 116 g 158 t
ORIGIN
Alignment Scores:
Pred. No.: 8,76e-95 Length: 538
Score: 855.00 Matches: 163
Percent Similarity: 98.20% Conservative: 1
Best Local Similarity: 97.60% Mismatches: 3
Query Match: 34.84% Indels: 0
DB: 10 Gaps: 0
US-09-934-066-2 (1-466) x AV555212 (1-538)
QY 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 38 ATGCTCTAGTCTCTCTGTCACCTTTCAGATCTTGGGTGCTCTTTTCTTTGCTTATCTTC 97
QY 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAspValGluSerSerAspLys 40
Db 98 TCAGCTGAGTCCCGCAAAACCCCAATTCCTGACCATTAATGATGTTGAATCTACGACAA 157
QY 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db 158 AGTCAAAAGGACACACGATGGGCTGTTTAGTTGCTGGATCAAAATGAATATTAACTAC 217
QY 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp 80
Db 218 AGGCATCAGCTGACATATGCCACGCTATCAGATATCCGAAAGGCGGCTTTAAAGAT 277
QY 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 278 GAAACATCATCTGTTTATGATGATATCGCGTTCCTCGGAGAAATCCTAGGCCT 337
QY 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 338 GGAGTTATCATTAATAACAGATGCGAAGATGTTTATAAAGGAGTTCCTAAGGACTAC 397
QY 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 398 ACTAAAGAGCTGTTAATGTTCAAAACTTCTACAATGTGTTACTTGAATGAAGCGGC 457
QY 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db 458 GTCACAGAGAGAAATGGCAAGTTGTGAAAAGTGGCCCTTAATGATATATCTTCACTAT 517
QY 161 TyrAlaAspHisGlyAlaPro 167
Db 518 TATGCTGACCATGGAGCTCCT 538
RESULT 15
AV562667 535 bp mRNA linear EST 07-SEP-2000
LOCUS Arabidopsis thaliana green siliques Columbia Arabidopsis
DEFINITION thaliana cDNA clone S0173h03f 3', mRNA sequence.
ACCESSION AV562667
VERSION AV562667.1 GI:8734093
KEYWORDS EST.

```

SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Euclidytedons; Core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 535)
 AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7, 175-180 (2000)
 MEDLINE 20363093
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 FEATURES
 source
 1..535
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="SQ173h03F"
 /clone_lib="Arabidopsis thaliana green siliques Columbia"
 /tissue_type="green siliques"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"
 BASE COUNT 170 a 92 c 114 g 159 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,36e-94 Length: 535
 Score: 848.00 Matches: 162
 Percent Similarity: 98.79% Conservative: 1
 Best Local Similarity: 98.18% Mismatches: 2
 Query Match: 34.56% Indels: 0
 DB: 10 Gaps: 0
 US-09-934-066-2 (1-466) x AV562667 (1-535)
 QY 1 MetSerSerProLeuGlyHisPheClnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
 DB 40 ATGTTTACTCTCTTGGTCACCTTCAGATCTCTGTTCTTATGCTTTGCTTATCTTC 99
 QY 21 SerAlaGluSerArgLysThrClnLeuLeuAsnAspValGluSerSerAspLys 40
 DB 100 TCAGCTGAGTCCCGCAAAACCAATTGCTGAACGATATGTTGAATCTAGCCACAAG 159
 QY 41 SerAlaLysGlyThrArgTIPAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
 DB 160 AGTGCAAAAGGCACACGATGGCTGTTTACTTGTGTCGATCAATGAATATTATAACTAC 219
 QY 61 ArgHisGlnAlaAspIleCysHisAlaTyrClnIleLeuArgLysGlyLeuLysAsp 80
 DB 220 AGGCATCAGGTGACATATGCCCGGTATCAGATCTCCGAAAGCGGTTTAAAGAT 279
 QY 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
 DB 280 GAAACATCATGTGTTTATGATGATGATATCGGTTTCTCGAGAAATCCTAGGCCT 339
 QY 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
 DB 340 GGAGTTATCATTAATAAACACGATGGAGAGATGTTTATAAAGGAGTCCCTAAGGACTAC 399
 QY 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
 DB 400 ACTAAAGAGCTGTTAATGTTTCAAAACCTTACAACTGTTTACTTGGAAATGAAAGTGC 459
 QY 141 ValThrGlyLysAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
 DB 460 GTCACAGGAGGAAATGGCAAAAGTTGTGAAAAGTGGTCTTATGATAATATCTTCATCTAA 519
 QY 161 TyrAlaAspHisGly 165

Db 520 TATGCTGACCATGGA 534
 Search completed: May 29, 2003, 19:45:27
 Job time : 1612 secs